



Edited by
Andy Neely

Business Performance Measurement

Unifying Theory and
Integrating Practice

Second Edition



CAMBRIDGE

CAMBRIDGE

www.cambridge.org/9780521855112

This page intentionally left blank

Business Performance Measurement

Drawing together contributions from leading thinkers around the world, this book reviews recent developments in the theory and practice of performance measurement and management. Significantly updated and modified from the first edition, the book includes ten new chapters that provide a comprehensive review of performance measurement from the perspectives of accounting, marketing, operations, public services and supply chain management. In addition to these functional analyses the book explores performance measurement frameworks and methodologies, practicalities and challenges, and enduring questions and issues. Edited by one of the world's leading experts on performance measurement and management, *Business Performance Measurement* will be of interest to graduate students, managers and researchers who wish to understand more about the latest developments in this rapidly changing field.

Andy Neely is Deputy Director of the ESRC/EPSRC AIM Research initiative, Professor of Operations Strategy and Performance at Cranfield School of Management, and Visiting Professor of Operations Management at London Business School.

Business Performance Measurement

Unifying theories and integrating practice

Second edition

Edited by

Andy Neely



CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521855112

© Cambridge University Press 2002, 2007

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2007

ISBN-13 978-0-511-37696-2 eBook (EBL)

ISBN-13 978-0-521-85511-2 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of urls for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>List of figures</i>	<i>page</i> ix
<i>List of tables</i>	xi
<i>List of boxes</i>	xii
<i>List of contributors</i>	xiii

Introduction	1
Andy Neely	

Part I Performance measurement – functional analyses and theoretical foundations **7**

1	Accounting performance measurement: a review of its purposes and practices David Otley	11
2	Measuring marketing performance: research, practice and challenges Bruce Clark	36
3	Measuring performance: the operations management perspective Andy Neely	64
4	Measuring performance: the supply chain management perspective Douglas Lambert and Michael Knemeyer	82
5	Finding performance: the new discipline in management Marshall Meyer	113

- 6 A conceptual and operational delineation of performance 125
Michel Lebas and Ken Euske

Part II Performance measurement – frameworks and methodologies 141

- 7 Performance measurement frameworks: a review 143
Andy Neely, Mike Kennerley and Chris Adams
- 8 Beyond budgeting to the adaptive organization 163
Jeremy Hope
- 9 Theoretical conditions for validity in accounting performance measurement 179
Hanne Nørreklit, Lennart Nørreklit and Falconer Mitchell
- 10 The validity of measurement frameworks: measurement theory 218
Stephen Pike and Goran Roos

Part III Performance measurement – practicalities and challenges 237

- 11 Choosing marketing dashboard metrics 239
Tim Ambler and John Roberts
- 12 Risk in performance measurement 261
Andrew Likierman
- 13 Measuring knowledge work 279
Rob Austin and Pat Larkey
- 14 Measuring innovation performance 304
Riitta Katila
- 15 Context-based measurement 318
Graham Clark and Antony Powell

Part IV Performance measurement in public services 337

- 16 Measuring the performance of England's primary school teachers: purposes, theories, problems and tensions 339
Andrew Brown
- 17 Police performance: sovereignty, discipline and governmentality 363
Paul Collier
- 18 The development of composite indicators to measure health care performance 383
Rowena Jacobs, Peter Smith and Maria Goddard
- 19 Perversity in public service performance measurement 408
Mike Pidd

Part V Performance measurement – emerging issues and enduring questions 431

- 20 Does pay for performance really motivate employees? 433
Margit Osterloh and Bruno Frey
- 21 Anomalies of measurement: when it works, but should not 449
Rob Austin and Jody Hoffer Gittel
- 22 Loosely coupled performance measurement systems 477
Thomas Ahrens and Chris Chapman
- Index* 492

Figures

1.1	The pyramid of financial ratios	<i>page</i> 14
2.1	A framework for marketing measurement	41
2.2	A contextual framework for marketing performance measurement (MPM) systems	51
3.1	The multiple dimensions of the five operations performance objectives	69
4.1	Supply chain management: integrating and managing business processes across the supply chain	83
4.2	Supply chain management framework: elements and key decisions	85
4.3	Supply chain network structure	86
4.4	Inventory flows within the supply chain	93
4.5	Annual inventory carrying costs compared to inventory turns for the manufacturer in table 4.1	94
4.6	Supply chain complexity	101
4.7	How customer relationship management affects economic value added	103
4.8	How supplier relationship management affects economic value added	104
4.9	Combined customer–supplier profitability analysis	105
4.10	How order fulfilment affects economic value added	107
4.11	Profit and market capitalization increase measured across four tiers of a supply chain	108
6.1	The performance tree	128
6.2	Performance is only relative	135
7.1	The performance measurement matrix	145
7.2	The SMART pyramid	146
7.3	The results–determinants framework	147
7.4	The input–process–output–outcome framework	147
7.5	The balanced scorecard	148

7.6	The European Foundation for Quality Management framework	149
7.7	The performance prism	155
8.1	Mintzberg's different views of strategy	166
9.1	Epistemic levels	201
9.2	The ontological and epistemological nature of accounting data	209
10.1	Stakeholder "claim" categorization	224
12.1	Risk and performance: three perspectives	265
12.2	Risk heat chart	270
14.1	Average citation pattern for 242 US biotechnology patents	309
14.2	Most innovative biotechnology firms in 1985–6 by citation-weighted patent measures	314
15.1	General breakdown structure numbering convention	320
15.2	Excerpt from fuel gauge breakdown structure	321
15.3	CbM construction	327
15.4	Pairing the terms in adjacent dimensions	329
15.5	Software application using CbM to measure activities	330
15.6	Effort in activities across the process and stage dimensions	331
15.7	Cost of change to software components	332
19.1	Deming's PDCA cycle	410
19.2	The cybernetic control metaphor	414
19.3	Ambiguity, uncertainty and non-canonical practices	416
21.1	Alternative models of performance measurement	470
22.1	Two context dimensions of performance measurement systems	479
22.2	Case company organizational structure	483
22.3	Different managers categorized the case company differently	489

Tables

2.1	Representative marketing activities and measures	<i>page</i> 42
4.1	How supply chain position affects inventory carrying cost	93
7.1	Case examples of organizations using the performance prism	157
8.1	Performance appraisal formula for a business unit at Groupe Bull in the mid-1990s	176
9.1	Activities involved in achieving specific goals	204
10.1	Comparison of proper measurement and indicators	220
10.2	Description of scales	231
14.1	Definitions of radicality of innovation	305
14.2	Results of the Poisson regression analysis predicting radicality of innovation	313
15.1	Example CbM lexicon	324
17.1	Transition within the triangle of sovereignty–discipline–punishment over three time periods	377
18.1	Summary of international examples of the use of composite indicators and key considerations	395
19.1	Hofstede on control	415
19.2	Root definitions	420
19.3	Dysfunctionalities and reasons for measurement	423

Boxes

11.1	Example of problems with ROI	<i>page</i> 249
12.1	The influence of SEC risk disclosure requirements	263
12.2	Economic profit at Michelin	266
12.3	The Sharpe ratio	267
12.4	Value at risk	271
12.5	Risk mitigation at Land Securities	273
12.6	Economic profit	274

Contributors

Chris Adams, co-author of “Performance measurement frameworks: a review”, is a Visiting Fellow at Cranfield School of Management, UK.

Thomas Ahrens, co-author of “Loosely coupled performance measurement systems”, is Senior Lecturer in Accounting and Finance at the London School of Economics, UK.

Tim Ambler, co-author of “Choosing marketing dashboard metrics”, is Senior Fellow, London Business School, UK.

Rob Austin, co-author of “Measuring knowledge work” and “Anomalies of measurement: when it works, but should not”, is Associate Professor, Harvard Business School, USA.

Andrew Brown, author of “Measuring the performance of England’s primary school teachers: purposes, theories, problems and tensions”, is at the University of Cambridge, UK.

Chris Chapman, co-author of “Loosely coupled performance measurement systems”, is University Reader in Accounting at the University of Oxford, UK.

Bruce Clark, author of “Measuring marketing performance: research, practice and challenges”, is an Associate Professor in the Marketing Group of the College of Business Administration, Northeastern University, USA.

Graham Clark, co-author of “Context-based measurement”, is at the University of York, UK.

Paul Collier, author of “Police performance: sovereignty, discipline and governmentality”, is a Senior Lecturer in the Department of Finance and Accounting at Monash University, USA, and a Fellow of the UK’s Advanced Institute for Management Research.

Ken Euske, co-author of “A conceptual and operational delineation of performance”, is Professor of Accounting, Naval Postgraduate School, Monterey, USA.

Bruno Frey, co-author of “Does pay for performance really motivate employees?”, is Research Director, Center for Research in Economics at University of Zurich, Switzerland.

Jody Hoffer Gittell, co-author of “Anomalies of measurement: when it works, but should not”, is Associate Professor, The Heller School for Social Policy and Management, Brandeis University.

Maria Goddard, co-author of “The development of composite indicators to measure health care performance”, is Assistant Director, Centre for Health Economics, University of York, UK.

Jeremy Hope, author of “Beyond budgeting to the adaptive organization”, is Founder and Research Program Director of BBRT, the Beyond Budgeting Round Table.

Rowena Jacobs, co-author of “The development of composite indicators to measure health care performance”, is Research Fellow, Centre for Health Economics, University of York, UK.

Riitta Katila, author of “Measuring innovation performance”, is Assistant Professor of Management Science and Engineering, Stanford University, USA.

Mike Kennerley, co-author of “Performance measurement frameworks: a review”, is a Research Fellow at Cranfield School of Management, UK.

Michael Knemeyer, co-author of “Measuring performance: the supply chain management perspective”, is Assistant Professor of Logistics, The Ohio State University, USA.

Douglas Lambert, co-author of “Measuring performance: the supply chain management perspective”, is Professor of Marketing and Logistics, The Ohio State University, USA.

Pat Larkey, co-author of “Measuring knowledge work”, is Professor of Public Policy and Decision Making, Carnegie Mellon University, USA.

Michel Lebas, co-author of “A conceptual and operational delineation of performance”, is Emeritus Professor Accounting and Management Control, Groupe HEC, France.

Andrew Likierman, author of “Risk in performance measurement”, is Professor of Management Practice, London Business School, UK.

Marshall Meyer, author of “Finding performance: the new discipline in management”, is Richard A. Sapp Professor and Professor of Management and Sociology, The Wharton School, USA.

Falconer Mitchell, co-author of “Theoretical conditions for validity in accounting performance measurement”, is Professor of Management Accounting, University of Edinburgh, UK.

Andy Neely, editor, author of “Performance measurement: the operations management perspective” and co-author of “Performance measurement frameworks: a review”, is Professor of Operations Strategy and Performance at Cranfield School of Management, UK, and Deputy Director of the UK’s Advanced Institute for Management (AIM) Research initiative, sponsored by the Economic and Social Research Council (ESRC) and the Engineering and Physical Sciences Research Council (EPSRC).

Hanne Nørreklit, co-author of “Theoretical conditions for validity in accounting performance measurement”, is Professor of Management Accounting and Control, Aarhus School of Business, Denmark.

Lennart Nørreklit, co-author of “Theoretical conditions for validity in accounting performance measurement”, is Professor at the Danish Centre for Philosophy and Science Studies, Aalborg University, Denmark.

Margit Osterloh, co-author of “Does pay for performance really motivate employees?”, is Professor for Business Management, Institute for Organisation and Administrative Science at University of Zurich, Switzerland.

David Otley, author of “Accounting performance measurement: a review of its purposes and practices”, is Professor of Accounting and Management, Lancaster University Management School, UK.

Mike Pidd, author of “Perversity in public service performance measurement”, is Professor of Management Science, Lancaster University Management School, UK, and a Fellow of the UK’s Advanced Institute for Management Research.

Stephen Pike, co-author of “The validity of measurement frameworks: measurement theory”, is Director of Research at Intellectual Capital Services Ltd, UK.

Antony Powell, co-author of “Context-based measurement”, is Managing Director of YorkMetrics Ltd, a University of York spin-out company that

provides policy-relevant consultancy, applied research and teaching on the measurement and management of complex projects and processes. He is also an honorary visiting fellow at the University of York where he was previously a lecturer in management systems.

John Roberts, co-author of “Choosing marketing dashboard metrics”, is Professor of Marketing, London Business School, UK.

Goran Roos, co-author of “The validity of measurement frameworks: measurement theory”, is a Visiting Professor at Cranfield School of Management.

Peter Smith, co-author of “The development of composite indicators to measure health care performance”, is Professor of Health Economics, University of York, UK.

Introduction

Business performance measurement: unifying theories and integrating practice

Andy Neely

In July 2000 I wrote the preface to the first edition of this book, which read:

Performance measurement is on the agenda. New reports and articles on the topic have been appearing at a rate of one every five hours of every working day since 1994. A search of the World Wide Web reveals over 170,000 sites dedicated to it. In 1996, one new book on the subject appeared every two weeks in the US alone. Since 1994 Business Intelligence, a professional conference organising company based in the UK, has organised some 90 separate events on business performance measurement (BPM). Some 2,700 delegates from over 1,400 different firms attended these conferences. In terms of delegate fees alone, Business Intelligence has accrued over \$5 million. Add to this, the revenues received by other conference organisers, publishers, market research firms, software vendors and consultants and it is clear that business performance measurement is a multi-million dollar industry.

Like many multi-million dollar industries developments are rapid. Recent years have seen the introduction of new methods of measurement, such as activity-based costing, throughput accounting and shareholder value analysis. New measurement frameworks, most notably the balanced scorecard and the business excellence model, have taken the business community by storm. Data collected by the US research firm, Gartner, suggest that 70% of firms will be using balanced scorecards to measure business performance by the end of 2000. Other data, such as that collected by the US consulting firm Towers Perrin, indicate that the majority of firms have introduced their balanced scorecards during the last five years. Similar trends can be observed in the field of quality management. Self-assessment frameworks, such as those underpinning the Baldrige Award and European Foundation for Quality Management Award, have generated significant industrial interest and activity. Increasingly authors and commentators are

discussing the multiple roles of measurement. It is now recognised that measures allow managers to do far more than simply check progress. The behavioural consequences of measures are frequently discussed. The value of benchmarking and external comparisons is widely understood. The question of what data should be disclosed to external parties – especially shareholders – is actively debated. Organisations such as Skandia, the Swedish Insurance company, and Shell have begun producing supplements to their annual reports. Skandia's supplement discusses the value of the firm's intangible assets, while Shell's 1998 supplement, entitled *Profits and Principles – Does There Have to Be a Choice?*, describes the company's environmental performance. Other organisations, such as the UK's Cooperative Bank, have moved even further and structured their annual report around the "inclusive" framework, proposed by the Royal Society of Arts, Manufacturers and Commerce following their Tomorrow's Company Inquiry. The Cooperative Bank's "Partnership Report", for the 1997 financial year, reviews the bank's performance through the eyes of its seven partners: (i) shareholders, (ii) customers, (iii) staff and their families, (iv) suppliers, (v) local communities, (vi) national and international society and (vii) past and future generations of "Co-operators".

In the academic community people from a wide variety of different functional backgrounds are researching the topic of performance measurement. Experts in accounting, economics, human resource management, marketing, operations management, psychology and sociology are all exploring the subject. One of the major problems with the field is that they are all doing so independently. The accountants discuss their ideas with other accountants. The operations managers talk to other operations managers. Rarely is knowledge generated in one academic functional academic silo assimilated by another. Of course, the end result is a massive duplication of effort. In 1998 the first multi-disciplinary conference on performance measurement was held at Churchill College in Cambridge. Between them the 94 papers presented at the conference cited some 1,246 different books and articles. Of these, less than 10% were cited more than once and only 0.3% were cited more than five times. These figures are symptomatic of a field with vast richness, but unmanageable diversity.

A significant problem this gives is that there appears to be no agreement as to which are the important themes and theories in the field. Everyone writing about the topic has his/her own preferred references and freely cites them. While this diversity is appealing, it also hinders development, because it makes it almost impossible for generations of researchers to build upon the work of others. If the field of performance measurement is to develop as an academic discipline then it is essential that some boundaries are identified and some theoretical foundations agreed. It is hoped that this book will make some small contribution to facilitating this process.

The last six years have seen some significant changes. An updated version of the first paragraph today, writing as I am in September 2006, would read:

Performance measurement is on the agenda. The ISI Web of Knowledge lists 6,365 scientific publications on performance measurement for the time period 2001–5. This is equivalent to one new scientific paper appearing on the subject every seven hours of every working day. Add to this the practitioner and popular literature, and it is easy to see why a Google search now reveals over 50,000,000 websites dedicated to performance measurement. In December 2005 Amazon listed over 3,700 books on performance measurement, while the latest estimates suggest that there are close to 100 software vendors selling performance-reporting packages. Add to the software vendors' revenues the fees collected by conference organizers, publishers, market research firms and consultants, and it is clear that organizational performance measurement is a multimillion dollar industry.

These updated facts and figures tell only part of the story. In the last five years international interest in performance measurement has grown significantly. It is clear that other regions of the world, most notably the Middle East and Asia, are now just as interested in performance measurement as Western economies. Across the world, governments are requiring public services to develop and deploy more sophisticated performance measurement and management systems. And, in light of corporate governance scandals, requirements to disclose information are increasingly being imposed on organizations, by legislators and by the investment community alike.

Of course, topical subjects always attract members of the academic research community – either as critics or advocates. One of the joys of academia is the diversity of views that are held and expressed. In the preface to the first edition of this book I used data from the Performance Measurement Association (PMA) conference to illustrate this point, highlighting that at the PMA's first conference (held in 1998) the “94 papers presented... contained references to some 1,246 different books and articles. Of these, less than 10% were cited more than once and only 0.3% were cited more than five times.” Updating these figures provides an equally interesting picture. Recently I completed a citation analysis of scholarly publications in performance measurement (Neely, 2005). The headline finding of this study was that the 1,352 performance measurement papers analysed included 31,646 citations, covering 25,040 works and drawing on 16,697 different lead authors. In the entire data set twelve lead authors were cited between fifty and 100 times, 266 were cited between ten and forty-nine times and 11,929 (71.4 per cent) were cited only once.

The data illustrate that the field of performance measurement still exhibits massive diversity: that 1,352 journal articles can cite 25,040 different works suggests a rich and diverse intellectual heritage. My aim in this volume is to draw together some of this rich diversity. Deliberately, the book draws on authors from a variety of functional disciplines, all of whom are working in the field of performance measurement. Deliberately, it presents a variety of perspectives on performance measurement.

The book opens with a section on functional perspectives and theoretical foundations, in which eminent authors from the accounting and finance, marketing, operations management, supply chain and psychology fields offer their views on measurement.

The second section of the book is devoted to a review of frameworks and methodologies. There are numerous such frameworks and methodologies, each with different strengths and weaknesses. The second section explores these, and raises questions about the theoretical validity of some of these frameworks and methodologies from both philosophical and mathematical perspectives.

The third section of book investigates the practical applications and challenges of performance measurement. Once again, this section draws upon multiple themes and disciplines, and the applications of measurement systems in a wide variety of contexts – especially those that are difficult to deal with – are discussed.

The fourth section moves to the arena of public services specifically, exploring approaches to performance measurement in education, the police and health. It also contains a provocatively entitled chapter – “Perversity in public service performance measurement” – that looks into the challenges of developing and deploying measurement systems.

The fifth, and final, section offers some views on emerging issues and trends in performance measurement, including explorations of the link between pay and performance, the reasons why measurement sometimes works when it should not and some of my own thoughts on the emerging phenomenon of corporate performance measurement systems.

The breadth of the book means that it should appeal to a wide audience, encompassing measurement scholars and practitioners. The book deliberately draws on work being undertaken by a diverse group of researchers – diverse in the sense of both geographical location and functional persuasion. The resultant richness illustrates well the diversity inherent in the topic of performance measurement, but it is hoped that the text also offers a reasonable foundation on which future generations of researchers can draw.

By bringing together these widely varied contributions in a single volume it is hoped that at least a start has been made on the process of unifying theories and integrating practice in performance measurement.

Reference

Neely, A.D. (2005). The evolution of performance measurement research: developments in the last decade and a research agenda for the next. *International Journal of Operations and Production Management*, 25(12), 1264–77.

PART I

Performance measurement – functional analyses and theoretical foundations

By its nature performance measurement is a diverse subject. Researchers with functional backgrounds as varied as accounting, operations management, marketing, finance, economics, psychology and sociology are all actively working in the field. As discussed in the introduction, this incredible diversity brings with it both challenges and opportunities. It results in a fascinating richness, but also makes it extremely difficult for each generations of researchers to build upon one another's work. A significant barrier stems from the fact that, traditionally, the way academic careers develop is through functional specialization. Accountants talk to accountants. Operations managers meet with operations managers. Marketing specialists network with other marketing specialists. The result is deep and rich streams of functionally specialized research, often with limited cross-fertilization. The aim of this first section is to begin the process of redressing this shortcoming by drawing together several functionally based reviews of performance measurement.

The section contains six chapters. The first, by David Otley, reviews measurement from an accounting and finance perspective and explores the different roles of measurement. Otley argues that the accounting community implicitly recognizes that measurement systems have three fundamentally different roles in organizations. First, they provide a tool for financial management. Second, they provide an objective for overall business performance. Third, they provide a means of motivation and control. A key theme in Otley's contribution is that far too often academics and practitioners do not recognize these three different roles, and the result can be significant confusion, especially when a measurement system designed to fulfil one role is used for another.

The second contribution comes from Bruce Clark, who provides an extensive review of marketing performance measurement. In his chapter Clark explores the theoretical and practical challenges of measuring marketing performance. He argues that this has become a particularly important issue in the past several years, and is complicated by inconsistent definition, varying organizational roles and the lagged effects of many marketing tools on customer behaviour. The chapter reviews these challenges, and then presents the means by which various scholars and practitioners have addressed them, with a special emphasis on research and practice since 2000. The chapter concludes with several issues and challenges to move marketing performance measurement forward.

The third contribution is based on the operations management perspective and is provided by Andy Neely. He explores the evolution of performance measurement research in operations, examining three different time periods. The first, prior to the 1980s, was a difficult period for the operations management community, as it was largely marginalized from the mainstream academy. The second, 1980–2005, saw significant developments in the field of operations management, stimulated partly by the emergence of Japan as a major economic power, which in turn resulted in a surge of interest in how better to manage operations. The third phase, since 2005, explores the question of “What next?” for performance measurement research in operations.

In the fourth chapter Douglas Lambert and Michael Knemeyer explore the issue of supply chain performance measurement. This contribution builds on some of the themes raised in Andy Neely’s chapter, extending the analysis to encompass the entire supply chain. Lambert and Knemeyer argue that many of the so-called supply chain metrics currently in use are no more than logistics measures, and call for a more holistic approach to supply chain measurement.

The fifth contribution is provided by Marshall Meyer, who argues that performance has the potential to become a new management discipline. Starting with the question “What is performance?”, Meyer argues that performance measurement, if used correctly, offers the potential for managers to understand which of the activities undertaken generate revenues that exceed costs. Developing this theme, he introduces the notion of activity-based revenue as a measurement methodology and illustrates how this approach has the potential to overcome some of the shortcomings encountered in the measurement systems used by organizations today.

The sixth and final chapter in this first part comes from Michel Lebas and Ken Euske, who ask explicitly: “What is performance?” Lebas and Euske

describe performance as one of those “suitcase words in which everyone places the concepts that suit them, letting the context take care of the definition”. They argue that this is one of the reasons why it is so difficult to develop theories in the field, and suggest that performance should be equated with purposeful action taken today designed to produce meaningful results tomorrow. Building upon this theme, Lebas and Euske then develop nine propositions designed to illustrate how performance can best be defined and understood through causal models shared by organizational decision makers.

1

Accounting performance measurement: a review of its purposes and practices

David Otley

Introduction

Accounting measures of performance have been the traditional mainstay of quantitative approaches to organizational performance measurement. However, over the past two decades a great deal of attention has been paid to the development and use of non-financial measures of performance that can be used both to motivate and report on the performance of business (and other) organizations. The impetus for such developments has come from the bottom and the top of the organization alike. Much performance management at the operational level is carried out using specific indicators of performance that are usually not measured in financial terms. At the most senior levels, although financial performance is inevitably a major consideration, there has been an increasing recognition that other important factors in the effective running of the organization cannot be well captured by such measures. Thus, non-financial performance measures have undergone significant development, to the relative neglect of the development of improved financial measures. However, the recent publicity surrounding the marketing of economic value added (EVA[®]) as an overall measure of company performance by management consultants Stern Stewart can be seen as a sign of a new emphasis on the financial aspects of performance.

The purpose of this chapter is to review the roles and functions of financial measures of organizational performance, and to outline the major features of their development, particularly in the last thirty years. It will be argued that there are three different major functions for the use of financial performance measures, and that, although these uses overlap to some extent, major confusion can be caused by applying measures developed for one purpose to a different use. The three main functions involved are as follows.

- (1) The use of financial measures of performance as a tool of financial management. Here the focus is on the functional specialism of finance and financial management. This is concerned with the efficient

provision and use of financial resources to support the wider aims of the organization, and to manage the effective and efficient operation of the finance function.

- (2) The role of financial performance as a major objective of a business organization. Here some overarching financial performance measure, such as profit, return on investment (ROI) or EVA[®], is used to signify the achievement of an important (perhaps the *most* important) organizational objective.
- (3) The function of financial performance measures as a mechanism for motivation and control within the organization. Here the financial information provides a “window” into the organization by which specific operations are managed through the codification of their inputs and outputs in financial terms.

Clearly, there is some overlap between these different functions. Efficient financial management is a component of efficient overall management, but it does not subsume the latter. Performance may be managed, in part, by the transmission of corporate objectives (in financial form) downwards as part of the process of strategy implementation, but other mechanisms and measures are generally required to supplement such financial measures. Financial measures may provide substantial insight into the overall impact of operational activities, but other, more specific, measures are generally needed to understand fully and manage the “drivers” of performance. This chapter will therefore first consider each of these major functions independently, and then examine the linkages between them.

What follows is by no means a comprehensive review of the uses to which financial performance measures have been put over the past thirty years. Rather, it is a brief report of the highlights of this area that attempts to draw out the lessons that have been learned, and that attempts to limit the confusion that can be caused by not recognizing the different functions involved.

A tool of financial management

Any organization, whether public or private, has to live within financial constraints and to deliver perceived value for money to its stakeholders. The role of the finance function is to manage the financial resources of the organization, and to ensure that the financial constraints it faces are not breached. Failure to do this will lead to financial distress, and ultimately, for many organizations, financial failure or bankruptcy.

Thus, financial planning and control is an essential part of the overall management process. Establishment of precisely what the financial constraints are and how the proposed operating plans will impact upon them are a central part of the finance function. This is generally undertaken by the development of financial plans¹ that outline the financial outcomes that are necessary for the organization to meet its commitments. Financial control can be seen as the process by which such plans are monitored and necessary corrective action proposed when significant deviations are detected.

There are three main areas of focus for financial plans. Most basically, cash flow planning is required to ensure that cash is available to meet the payments the organization is obliged to meet. Failure to manage cash flows will result in technical insolvency (the inability to meet payments when they are legally required to be made). For business organizations, the second area requiring attention is profitability, or the need to acquire resources (usually from revenues acquired by selling goods and services) at a greater rate than using them (usually represented by the costs of making payments to suppliers, employees and others). Although, over the life of an enterprise, total net cash flow and total profit are essentially equal, this can mask the fact that, in the short term, they can be very different.² Indeed, one of the major causes of failure for new small business enterprises is not that they are unprofitable but that the growth of profitable activity has outstripped the cash necessary to resource it. The major difference between profit and cash flow is in the acquisition of capital assets (that is, items of equipment that are bought and paid for immediately, but that have likely benefits stretching over a considerable future period) and timing differences between payments and receipts (requiring the provision of working capital). This highlights the third area of focus, namely on assets and the provision of finance for their purchase. In accounting terms, the focus of attention is on the balance sheet, rather than the profit and loss account or the cash flow statement.

In overall terms, financial management therefore focuses on both the acquisition of financial resources on as favourable terms as possible and the utilization of the assets that those financial resources have been used to purchase, as well as looking at the interaction between these two activities.

¹ Such financial plans are often referred to as budgets and are widely used as a means of management control. However, this use is more concerned with management control than financial control, and will be discussed in later sections.

² If “clean surplus” accounting is used, total net cash flow and total profit are identical, in aggregate.

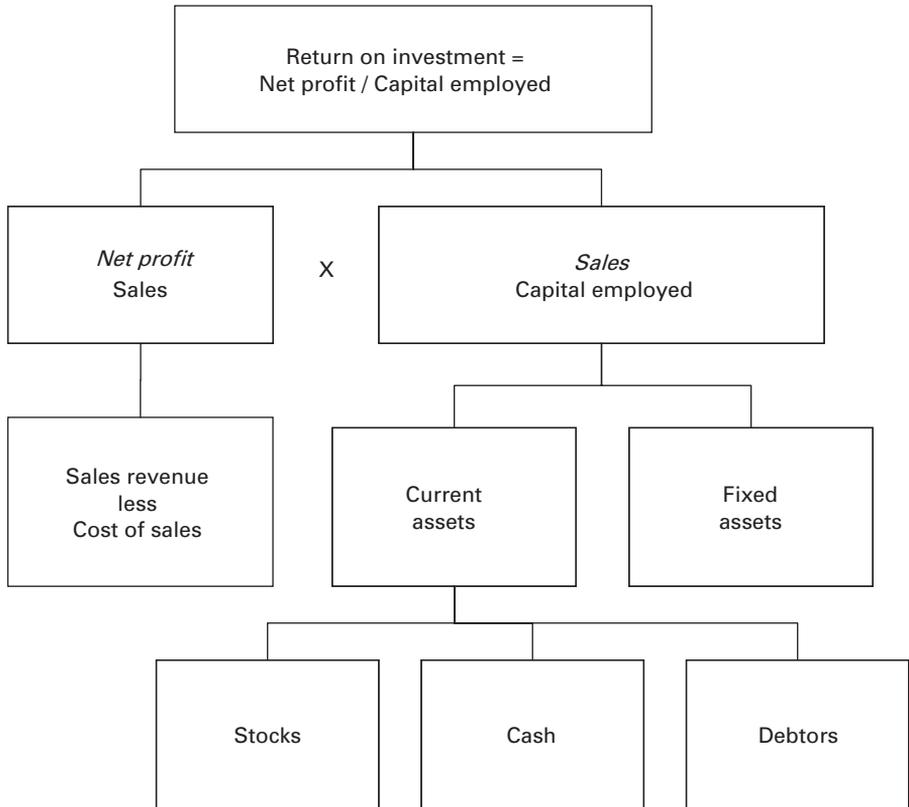


Figure 1.1: The pyramid of financial ratios

The single most powerful tool of reporting on these matters is the so-called “pyramid of ratios”.

The apex of the pyramid of ratios (see figure 1.1) is an overall measure of profitability that divides profit by the assets used in generating that profit, namely return on capital employed. Traditionally, this is broken down into two major secondary ratios, namely the profit margin on sales and the capital turnover. Clearly, return on capital employed is equal to the product of these two items. Each of the secondary ratios can be broken down into tertiary ratios based on the fact that profit is equal to sales revenue less cost of sales, and capital employed can be split into fixed assets (long-term) and current assets (short-term). However, it is evident that the concept becomes more strained the further down the pyramid one proceeds, and, although the pyramid provides a clear connection between the values of each of its component ratios, a more focused approach can be more beneficial than attempts to create a totally integrated “pyramid”.

This can be provided by considering the purpose of calculating each ratio. Thus, if the concern is with cash flows and liquidity, a range of ratios based on working capital are appropriate. Thus, five key ratios are commonly calculated, namely

- (1) current ratio, equal to current assets divided by current liabilities;
- (2) quick ratio (or acid test), equal to quick assets (current assets less inventories) divided by current liabilities;
- (3) inventory turnover period, equal to inventories divided by cost of sales, with the result being expressed in terms of days or months;
- (4) debtors to sales ratio, with the result again being expressed as an average collection period; and
- (5) creditors to purchases ratio, again expressed as the average payment period.

Each of these ratios addresses a different aspect of the cash collection and payment cycle. There are conventional values for each of these ratios (for example, the current ratio often has a standard value of 2.0 mentioned, although this has fallen substantially in recent years because of improvements in the techniques of working capital management, and the quick ratio a value of 1.0), but in fact these values vary widely across firms and industries. More generally helpful is a comparison with industry norms, and an examination of the changes in the values of these ratios over time that will assist in the assessment of whether any financial difficulties may be arising.

If the concern is more with longer-term profitability than with short-term cash flows, a different set of ratios may be appropriate. Profit to sales ratios can be calculated (although different ratios can be calculated depending whether profit is measured before or after interest payments and taxation); value added (sales revenues less the cost of bought-in supplies) ratios are also used to give insight into operational efficiencies. A general principle is that each part of the ratio should be relevant to the audience being addressed, and that the overall ratio should reflect the specific user of the information it provides.

Finally, if it is wished to consider the raising of capital as well as its uses, a further set of ratios based on financial structure can be calculated. For example, the ratio of debt to equity capital (gearing or leverage) is an indication of the risk associated with a company's equity earnings (because debt interest is deducted from profit before profit distributable to shareholders is obtained). It is often stated that fixed assets should be funded from capital raised on a long-term basis, whilst working capital should fund only short-term needs. Again, this may seem to be a logical and prudent rule of thumb,

but it is necessary to be aware that some very successful companies flout this rule to a very considerable extent. For example, most supermarket chains fund their stores (fixed assets) out of working capital because they sell their inventories for cash several times before they have to pay for them (i.e. typical inventory turnover is three weeks, whereas it is not uncommon for credit to be granted for three months by their suppliers). Thus, the values of these ratios indicate the potential riskiness of such an arrangement, but this does not necessarily preclude such a financial strategy being adopted.

It is of note that the overall return on investment ratio can be calculated in a variety of different ways. For example, return (profit) may be before or after payment of debt interest. Capital employed may be measured as total capital employed in the business, or just as the equity (shareholders') capital alone. Which measure is appropriate depends upon the use to which the ratio is being put. If the focus of interest is the efficient use of financial resources by the firm as an entity, then profit before interest and taxation (PBIT) may appropriately be divided by total capital employed. If the area of interest is the use of shareholders' capital, then the return attributable to shareholders (i.e. profit after interest and taxes (PAIT)) divided by equity capital alone may be the more meaningful measure.

There is, therefore, no definitive set of financial ratios that can be said to measure the performance of a business entity. Rather, a set of measures can be devised to assess different aspects of financial performance from different perspectives. Although some of these measures can be calculated externally, being derived from annual financial reports, and can be used to assess the same aspect of financial performance across different companies, care needs to be taken to ensure that the same accounting principles have been used to produce the accounting numbers in each case. As company directors are well aware that such analyses may be performed, it is not uncommon for "window dressing" to occur so that acceptable results are reported. A considerable amount of such manipulation is possible within generally acceptable accounting principles (GAAPs), although occasionally the manipulation strays into the realm of more "creative accounting", which may fall foul of the auditors.

Such "creative accounting" draws attention to the interface between management accounting information (which is intended to be useful in internal decision making and control) and financial accounting (which is a major mechanism by which external stakeholders, most notably shareholders, may hold managers accountable for their stewardship). Recent financial scandals, such as Enron and WorldCom, have highlighted the possibilities for financial

accounting reports to give misleading signals whilst still, arguably, following GAAP. Part of the cause has been the “rules-based” approach of US financial reporting, in contrast to the “principles-based” approach of the UK regime. But this is not the whole story, for the United Kingdom has not been free of such scandals of its own in the not too distant past (e.g. *Pretty Polly*, *Maxwell*). As judgement is required to a significant extent in constructing financial reports, there will always be scope for it to be exercised to produce a misleading picture, given the huge pressures on organizations and their managers to be seen to be performing well.

One result of the reforms that have followed these scandals has been a greater emphasis on operating information. This is clear evidence that financial numbers alone are insufficient to reveal the overall financial condition of an enterprise. In addition, legislation such as the Sarbanes–Oxley Act in the United States has required a much greater disclosure of the potential risks surrounding an enterprise, reflected internally by a much greater emphasis on risk management and the maintenance of risk registers.³

However, financial ratios do allow internal financial managers to keep track of a company’s financial performance (perhaps in comparison with that of its major competitors), and to adjust the activities of the organization, both operating and financial, so as to stay within acceptable bounds. A virtuous circle can be constructed whereby net cash inflows are sufficient to pay adequate returns to financiers and contribute towards new investment; given sound profitability, the financiers will usually then be willing to make additional investment to finance growth and expansion beyond that possible with purely internal finance. Conversely, a vicious cycle can develop when inadequate cash flows preclude adequate new investment, causing a decline in profitability, and so the organization becomes unable to sustain itself.

From this perspective, the role of financial performance measurement is to help keep the organization on a financial track that is “straight and narrow”. The performance measures are used primarily by financial specialists, and the action taken as a result of such analysis may also be exclusively financial (e.g. raising more capital to ensure that debts can be paid on time). Nevertheless, it is also clear that financial problems because of deficiencies in other areas of business operations may also come to light. In this case, the ratios can provide the finance director with the information necessary to convince other managers that operating action needs to be taken in order

³ Risk and its association to performance measurement is discussed at length by Andrew Likierman in chapter 12.

to avoid financial distress. However, the primary role served by this type of performance measurement lies within the province of the finance function, and is concerned with the effective and efficient use of financial resources. But it should be stressed that the finance function serves a boundary role; it is an intermediary between the internal operations of an organization and the key external stakeholders who provide the necessary financial resources to keep the organization viable.

An overall business objective

The second major role of accounting performance measures is connected with the financial objectives of the business. In particular, measures are addressed to meeting the needs of the external suppliers of capital, both debt and equity. It is this need that external financial reporting addresses. An organization's annual report and financial accounts are produced primarily for the shareholders, although some use may be made of them by bankers and other providers of debt capital. In some ways, these external financial reports can be seen as mirroring the internal measures and ratios discussed above, in that they cover the same three main areas of cash flow (rather obliquely), operating profit and asset values. In addition, the two differing foci of the performance of the business entity (financed by both debt and equity capital) and the return to its shareholders (i.e. the return on equity capital alone) are also apparent. However, by far the most attention is focused on reporting to shareholders.

The whole area of external financial reporting, in particular, and the debate surrounding corporate governance, more generally, is structured around the usefulness of audited financial statements (and other mandatory disclosures) to shareholders. At one level, this is captured by the agency theory formulation, whereby owners (shareholders) seek to control managers but are restricted in their ability to do this because they possess much less detailed information than the managers. Mandatory accounting statements represent one means of attempting to redress this balance by providing shareholders with an annual, externally audited review of the financial outcomes associated with the business activities undertaken. This is very much of a "backstop" position, and active investors (e.g. institutional shareholders) generally seek to obtain more frequent and prospective information than financial reports can provide. However, the acquisition of prospective information is restricted by the need to make all such

information public, in order to preserve an equitable trading market in which all players have similar access to information. The whole area of “insider trading” and the legislation governing stock market operations is an example of the complexity of the rules needed to preserve such an open market. Thus, this brief review will restrict itself to the information provided by annual financial accounts to the shareholders of a business, and the measures of performance that are used in this respect.

The legal constitution of shareholder-owned enterprises puts the shareholders in the position of being the residual owners of any financial benefits that the organization may create through its activities. The profit and loss account eventually arrives at a figure of profit attributable to shareholders, after the deduction of all expenses including debt interest and taxes – PAIT. Other parties may have a legal right to certain fixed payments (e.g. supplier invoice payments, employees’ wages, etc.) but any excess over these expenses represents profit attributable to shareholders, without any upper limit. Generally, this profit will be partly distributed in the form of a dividend, and partly retained in the business (retained earnings) to finance future expansion. If an organization fails to make a profit, dividends may still be paid out of previously retained earnings, but ultimately this will become exhausted and the business will become bankrupt. In such a case, it is likely that the shares will be valueless, and the shareholders will lose their investment, up to the amount they invested. There is no provision for the recovery of any further losses from shareholders (i.e. their liability is limited to the amount they paid for their shares).

Thus, earnings (profit) are the central performance indicator for shareholders. A very common measure of performance is EPS (earnings per share), which divides total annual earnings by the number of shares issued. Earnings essentially represent the (cum dividend) increase in the accounting book value of the company due to its previous year’s activities. However, the share price of a company depends not only on the performance it achieved in the past but also on expectations of its future prospects. In technical terms, the share price “impounds” such information and conceptually represents both the historical value of the assets it possesses and the expectation of future performance, discounted by an appropriate time-value of money. The results of these future expectations is illustrated in the commonly calculated price/earnings (P/E) ratio, which divides the current share price by the last reported earnings figure. A high value of this ratio indicates an expectation of a high level of growth in future earnings, a low value an expectation of stability or even a decline.

Annual reported earnings thus represent only one component of the return to shareholders, and one often seen as being of only secondary importance. More formally, in any period of ownership the return to a shareholder comprises the dividends received plus the increase in share price (or minus the decrease in share price) that has taken place during the period, divided by the initial share price. By way of a practical example, the average growth in share price during the 1990s on the UK stock market was somewhat in excess of 15 per cent per annum, whereas dividends have been paid at a rate of around 3 to 4 per cent per annum. Thus, the bulk of the return to shareholders was in the form of capital growth rather than dividend payments. Further more, a period's dividend payment is related only loosely to the earnings in that period. Finally, the computation of actual returns to shareholders requires no accounting information whatsoever, being comprised of cash dividend payments and the change in the market price of the shares.

However, the experience of the first few years of the current century has been very different. Expected returns on equity have been reduced to less than a half of those of the previous decade (say 5 to 6 per cent). Although dividends have been reduced slightly, they now comprise a much more substantial part of the expected return. This change in economic expectations has had repercussions far beyond the stock market. The problems that individuals have experienced with endowment insurance policies not producing their expected returns, and part of the so-called pensions crisis, both stem in large measure from the reduced expectations of future growth now held.

What is a reasonable rate of return that may be expected by shareholders in a particular business? This question can be answered by only reference to past experience, and only in average terms. Essentially, the computation that is being performed is an assessment of the opportunity cost of capital to the investor (i.e. what return might a shareholder expect if he/she had invested in alternative, but similar, investments?). The capital asset pricing model (CAPM) has been a popular method of making this assessment, and concludes that the return that can be expected depends upon (a) the risk-free rate of return that can be obtained from investing in an interest-bearing investment such as a government bond, and (b) the riskiness of the particular investment being considered. This riskiness (represented by the coefficient beta in the model) is assessed by comparing the sensitivity of the returns from the particular investment with the returns from the market portfolio (i.e. a composite of all available investments, such as a stock market tracking fund). The relationship is assumed to be linear in form, so knowledge of a firm-specific beta and the risk-free and market portfolio expected returns allows an estimate of

the cost of a company's equity capital to be made. This provides a benchmark against which future returns can be assessed. If returns in excess of this benchmark are expected, the share price is likely to rise such that new investors will obtain a return exactly in line with the benchmark figure.

However, most commonly used measures of company performance do not match this model. Earnings are the fundamental component of many performance measures. To the extent to which such earnings-based performance measures are assumed to capture information about the values of a business, these measures implicitly assume that past earnings are a good predictor of future returns and are thus associated with share price. As future values are much more dependent upon expectations of future performance, it is not necessarily the case that a historical measure of past performance is likely to be strongly associated with share price. Moreover, the current share price of a company already impounds all the publicly available information (and possibly much private information, as well) about its future prospects. The market has, therefore, already taken into account all such information in setting the current share price. In a world of perfect information, the past history of company performance is irrelevant to predicting future share price movements.

Accounting measures of performance are restricted largely to providing confirmatory evidence that the beliefs of investors concerning current earnings are based on auditable "fact". But it must also be recognized that the calculation of accounting earnings is a matter of judgement as well as fact. For example, a charge representing the depreciation in value of capital assets forms a major cost item in the accounts of most companies. However, this requires an assessment to be made of the expected future life of these assets, and their likely residual value at the end of this life. Clearly, this requires the exercise of a considerable degree of judgement, and different accountants might well reach different conclusions as to the amount of profit to be reported. Less legitimately, the whole arena of "creative accounting" indicates the lengths to which accounting judgements can be stretched in the cause of reporting profit figures that are helpful to directors and others.

In an increasingly unpredictable world, less reliance can be placed on the history of past performance as a basis for predicting future performance. Indeed, to the extent that future earnings streams and value creation have become less predictable, asset values will tend to be reduced to reflect this greater uncertainty. Although accounting may provide a reasonable account of what has happened in the past, it is perhaps much less useful than ever before in forming a basis for assessing what may happen in the future.

Thus, paradoxically, even if the delivery of returns to shareholders is seen as the overall aim and objective of a business enterprise, reported accounting earnings provide only a weak surrogate for overall shareholder returns. There is a considerable body of empirical literature that demonstrates the relatively low level of correlation between reported profitability and share price movements.⁴ But, even in the absence of such evidence, it is clear on conceptual grounds that no such relationship is likely to be strong. In terms of assessing performance from an investor's perspective, accounting measures provide only background and confirmatory evidence. Even economic value added, which will be discussed in detail in the following sections, is essentially an accounting-based performance measure, and, as such, cannot be expected to do more than imperfectly mirror shareholder returns.

A mechanism for motivation and control

The third major function of accounting performance measurement lies in its internal use as a means of motivating and controlling the activities of managers so that they concentrate on increasing the overall value of the business, or, at least, the value attributable to the shareholders. In short, the role of managers is often presented as “increasing shareholder value”. Even if this is accepted as the overriding objective of the business, there is a complicated chain of means–end relationships that now needs to be considered. That is, how can shareholder value be increased?

At the first level of analysis, controllable aspects of performance can be partially captured in accounting performance measures, both earnings and balance sheet values. Here, the accounting information is not just concerned with financial performance but, rather, uses financial indicators to represent the underlying activities that are being managed. In an organization of any size or complexity, there is a need to be able to represent a variety of different activities in terms of a common language or unit of measurement. Accounting provides such a common language, so that the impact of very different activities can be aggregated into overall measures such as sales revenue, costs and profitability.

At the next level of analysis, it may be realized that measures of outcomes are an insufficient mechanism for controlling performance. What is required

⁴ See Lev (1989) for a summary of evidence from the first two decades of “market-based” accounting research.

in addition are measures that represent the “drivers” of performance – that is, those activities that it is believed are necessary to be undertaken so that desired outcomes (financial and otherwise) are attained. At this level, accounting measures alone become inadequate, and over the past ten years alternative approaches, such as the balanced scorecard, have been developed to supplement solely accounting measures of performance.⁵

The two sets of approaches based on these differing models are likely to be complementary, but the development of performance measurement has tended to divorce them. We shall, therefore, first consider the development of accounting-based performance management techniques, and then go on to review the wider approaches that have been developed more recently.

Accounting approaches to control

The basic accounting approach to motivation and control is to divide an organization into “responsibility centres”. These are organizational units that are as self-contained as possible, and that are responsible for defined aspects of performance. At the highest level, these are defined as investment centres, where managers have responsibility both for investing in business assets and for using the assets entrusted to them effectively. A typical performance measure for an investment centre manager would be return on capital employed, as this involves both profit and asset value components. At a lower level, profit centres are defined. Here managers are responsible for generating sales revenue and for managing the costs involved in production or service delivery. Thus, profit is an appropriate performance measure. Finally, the lowest level of responsibility is the cost centre, where the results of the units activity cannot be assessed in terms of revenue earned, and managers are held responsible (in accounting terms) only for costs.⁶ Clearly, in performance management terms, cost centres need to have other (non-accounting) measures associated with them in order to capture the outputs that result from expenditure on inputs.

To operate control based primarily on accounting measures requires profit or investment centres to be established. Indeed, there has been a

⁵ See chapter 7 for further information on the balanced scorecard and other measurement frameworks.

⁶ A further form of responsibility centre, the revenue centre, is sometimes used. This is where the unit generates sales revenues, but it is responsible only for the marketing costs, not the costs of producing the products sold. Here the net revenue figure can be used as a measure of the success of the unit.

tendency to create “pseudo-profit centres”, where revenues are somewhat artificially attributed to responsibility centres in order to gain the advantages associated with control of profit centres. These advantages are primarily those of having only to consider accounting measures of performance, expressed as an overall profit measure and its components. In particular, if a profit centre is indeed generating profits, it can potentially be left alone to continue the good work, and control can be exercised in a relatively decentralized manner. However, to construct profit statements for an organizational unit requires revenues as well as costs to be attributed to it. This is not an issue when products are sold to an external customer and sales revenues are generated, but it is more problematic when intermediate products are transferred internally within a larger organization or when, in the public sector, services may be provided at no cost to the immediate user.

Here a value has to be attributed to the transferred goods and services, the so-called “transfer price”. A great deal of attention has been paid in the accounting literature to the setting of transfer prices that will motivate managers to act in the interest of the overall organization whilst maximizing their own reported profit measure. This can be achieved under certain circumstances, but it is more common for transfer prices to generate more heat than light. If they are mis-set, there is considerable potential for managers to appear to be performing well in local terms but to be acting dysfunctionally from a more global perspective. An extreme example was the case of the motor car manufacturer that set transfer prices on a “full cost plus” basis – that is, each component plant, and the assembly plant, had transfer prices set on the basis of their full costs plus a percentage addition for their profit margin. Not surprisingly, all the units reported healthy profits; the only black spot was the marketing division, which reported heavy losses, as it was unable to sell the vehicles at anything like the costs that had been transferred to it. Clearly, in this case, the problem did not lie solely in the marketing area but also in the high production costs in all the other areas. Inappropriate setting of transfer prices, and the tendency to attempt to create profit centres when they do not really exist, are responsible for a great deal of dysfunctional activity.

Even when the transfer pricing issue has been satisfactorily dealt with, there is a further issue of motivation that can arise. In a profit centre, a manager can be targeted to improve his/her profit target. But, because the conventional calculation of profit excludes any assessment of the return required by the providers of equity capital, maximization of reported profits is not an appropriate objective. For example, it can be achieved by using

excessive investment in working capital to produce a low, but positive, rate of return. Conversely, in an investment centre, the use of return on investment as a performance measure can lead to underinvestment. For example, a manager currently achieving a high rate of return (say 30 per cent) may not wish to pursue a project yielding a lower rate of return (say 20 per cent) even though such a project may be desirable to a company that can raise capital at an even lower rate (say 15 per cent). Both these potentially dysfunctional motivational effects can be overcome by the use of residual income as a performance measure.

Residual income is defined as accounting profit less a charge for the equity capital used in its generation. That is,

$$\text{Residual income} = \text{Accounting income less (Capital employed} \times \text{Cost of capital \%)}^7$$

This overcomes the problems described above. Any project that increases residual income over the life of an asset is desirable; any project that decreases residual income is undesirable. In principle, the potential for dysfunctional motivation is removed, and residual income is thus a better overall measure of performance than either profit or return on investment. However, rather surprisingly, over the last thirty or more years since residual income was introduced in the academic literature it has been surprisingly little used in practice. But, recently, this lack of use has radically changed, especially in the United States. During the 1990s the US management consultants Stern Stewart introduced a performance measure that they named economic value added, which is conceptually identical to residual income, and they have marketed it very successfully as an overall performance measure for companies and as a device for measuring the performance of individual business units. They argue that all other performance measures in common use, including profit, return on investment and earnings per share, potentially create dysfunctional motivations for managers. To encourage managers to focus on creating shareholder value (rather than, for example, pursuing growth for its own sake, or because of the advantages growth can bring to the managers themselves) they argue that EVA[®] represents the one and only measure that unambiguously provides the appropriate motivation.

⁷ Note that this can be computed either by taking accounting income after interest charges and using the equity capital employed and its cost or, alternatively, by taking profit before interest and using the total capital employed and the weighted average cost of capital (WACC). The numerical result should be identical; however, the latter approach is usually more easily applied in practice because capital employed can be measured by valuing the assets involved.

To do this, they recommend that a considerable number of adjustments be made to the conventional financial accounts produced by companies. Most of these adjustments attempt to replace conventional financial reporting practices with recognition and measurement procedures that produce a more meaningful estimate of the capital committed to an enterprise by its investors. Stern Stewart claim to show, in a series of studies, that EVA[®] correlates more closely with share price than any other accounting measure. However, it still needs to be recognized that, even if this claim is substantiated, no historical measure of performance will be a perfect predictor of share price, in that much of the price of a share is determined by future expectations rather than past results.

Not only do Stern Stewart recommend the use of EVA[®] at the highest levels of the organization, they also strongly recommend that it be driven as far down the organization as possible, so that managers at every level are given the task of improving their reported EVA[®]. They also argue that managerial rewards should be closely matched to this performance measure. In such a way, they argue, managers will be motivated to improve shareholder value.⁸

Although much can undoubtedly be achieved in this way, there are also some limitations to the effectiveness of this approach. First, accounting performance measures for a single period cannot reflect accurately the impact of decisions that may have repercussions over several accounting periods. For example, it has been shown that capital investment decisions that have a positive net present value (NPV), and that should therefore add value to the firm, do not necessarily yield positive accounting profits (or returns on investment, or EVA[®]) in every period of the project's life. The only way to ensure such an outcome would be to value the assets concerned at the NPV of their future expected cash flows. Although this is acceptable in economic decision-making terms, it is not feasible from the viewpoint of reporting on performance, as such estimates would be excessively subjective. For example, a manager could improve on his reported performance merely by making slightly optimistic estimates of the outcome of future events. Second, even when multiple periods are considered, historical earnings represent the true growth in value of a business only if the assets it possesses are valued in terms of future expectations rather than historical attainments. That is, GAAPs

⁸ See Young and O'Byrne (2001) for an informed view of the benefits and limitations of using EVA[®] as a tool for value-based management. This book is a relatively balanced assessment of the EVA[®] performance management system, unlike many others that have an evangelical flavour.

would have to be cast aside and assets valued at the net present value of their expected future cash flows. At the very least, such an approach requires a great deal of subjective judgement on the part of managers, and is therefore open to significant manipulation. As a result, there are fundamental limitations as to what can be achieved by using historical accounting numbers to measure and assess managerial and organizational performance.

Performance drivers

The complementary approaches that have been developed move away from a concentration on accounting measures alone, and add consideration of a wider range of factors that are believed to drive future economic performance. The most popular of these approaches in the 1990s was the balanced scorecard approach, developed at the Harvard Business School (Kaplan and Norton, 1996). Although this will be discussed in some detail in this section, it should be recognized that other similar approaches exist, including the European Foundation for Quality Management (EFQM) scheme, which is in many ways similar to the Harvard approach. Moreover, these approaches are not new. The General Electric Company developed a set of performance measures for its departments in the 1950s that incorporated the following elements:

- short-term profitability,
- market share,
- productivity,
- product leadership,
- personnel development,
- employee attitudes,
- public responsibility, and
- balance between short-range objectives and long-range goals.

However, the balanced scorecard approach has a number of features that make it a good vehicle for structuring an array of performance measures. First, it makes an explicit link between the espoused strategies of an organization and the performance measures it uses to monitor and control strategy implementation. This key feature makes it very clear that there is not necessarily a universal set of performance measures that are appropriate for all organizations in all circumstance (as seems to be assumed in many accounting approaches), but that specific measures need to be devised for specific circumstances. Second, the four major areas in which performance

measures are to be devised (financial, customer, business process, and innovation and learning) closely match the main stakeholders of the organization (especially as the employees tend to be discussed in the fourth area – innovation and learning). It would not be difficult to extend the balanced scorecard approach into a more fully developed stakeholder model. Third, there is a clear attempt to model the main drivers of future performance, as each area requires the question “What must we do in order to satisfy the expectations of our stakeholders?” to be considered, and appropriate responses generated. Finally, the requirement that there be a limited number of performance measures in each area requires a focus on the “key success factors” that are believed to operate. This can help to compensate for the tendency to construct ever-increasing numbers of performance indicators. The difficulty in constructing a balanced scorecard lies not in generating enough performance measures but, rather, in selecting down to a very small number of centrally important measures.

In this formulation, the balanced scorecard uses measures of financial performance to ensure that the requirements of financiers are addressed. This closely matches the financial management use of accounting information, and may also incorporate some concept of an overall objective. Thus, it would seem that EVA[®] can quite appropriately be used as one of the financial measures in a balanced scorecard formulation. Interestingly, financial measures may appear in other areas. For example, the proportion of revenue generated by new products is cited as a learning and innovation measure. Clearly, a measure derived from financial components is being used to assess the long-run future prospects of a business unit. In a similar way, customer satisfaction may be assessed by repeat business, again measured by sales revenues. Such an approach perhaps gives greater insight into the development of appropriate accounting performance measures than the more universalistic approaches that accountants have tended to espouse.

Connections between the approaches

Although three major functions of accounting performance measures have been distinguished in the preceding sections, it is also common for any particular accounting measure to be used for more than one of these functions. For example, return on investment may be seen as the peak of the financial effectiveness pyramid of ratios, as a major business objective in its

own right and as a key performance target used to motivate and monitor operating performance.⁹

The most-studied accounting technique in this regard has been the process of budgetary control. Although a budget consists of a whole set of accounting numbers, the “bottom line” (i.e. either total costs or operating profit) forms a single performance measure in its own right. Two major functions of budgets have been distinguished. First, a budget can be used as a financial plan, utilized by the finance department to ensure that the organization stays within its operating constraints. Second, it is much more widely used in most organizations as a tool of overall management control. Here, budget targets are set for individual responsibility centres and their operating managers, so that operating performance can be monitored and controlled. The aggregate of all the responsibility centre budgets becomes the overall operating objective of the organization, expressed in financial terms. The budgeting literature is very clear that budgets can be used for these two – and other – major purposes within an organization. It is equally clear that a single budget system cannot serve all these diverse purposes equally well, and that decisions need to be made as to which purposes should be prioritized.

The most acute conflict is often between the two functions outlined above, where the same budget estimates are used both for financial planning and management control. This frequently results in neither purpose being adequately served. Financial planning estimates need to be “best estimates” of likely outcomes, or even conservative estimates given the unpleasant consequences of becoming unexpectedly illiquid. By contrast, management control is often best served by budget estimates being set as motivational targets that are “challenging, yet attainable”. In practice, such targets may frequently fail to be achieved, yet they may have served their purpose of motivating maximum managerial effort.

It is of interest to note that some of the more recent literature on budgetary control (see Bunce, Fraser and Woodcock, 1995) indicates a widespread dissatisfaction on the part of users of traditional budgetary control techniques, because they are seen to be failing as adequate control devices. This is partly because of the levels of uncertainty faced by organizations and the difficulties of making accurate forecasts of future events; the budget is often regarded as being out of date even before the budget period has begun.

⁹ Advocates of EVA[®] would no doubt argue that EVA[®] is an even better measure to use for these purposes.

In such organizations, the primary role of budgeting is reverting to that of financial planning, with management control being assisted by a variety of measures of operating performance that are non-financial in nature, perhaps organized using a balanced scorecard framework. In this context, it is also of interest that Stern Stewart's views of budgeting are quite clear; it is a useful financial planning technique, but should not be used as a basis for issuing incentives and rewards.

This issue has been well documented by the "beyond budgeting" (BB) movement (see Hope and Fraser, 2003). Many of the (increasing) problems experienced by organizations attempting to manage their budgetary control procedures have been laid at the door of the budget representing a "fixed" (preset and unchanging) target in a world of constant change and uncertainty. Thus, budgets can become outdated during the budgetary year, or even before it begins.

The weaknesses identified by the BB movement are very similar to a list, also of twelve items, subsequently outlined in a report by Neely, Sutcliffe and Heyns (2001). Drawn primarily from the practitioner literature, they maintained that the twelve most-cited weaknesses of budgetary control (which can also be applied more generally to management control) are:

- budgets constrain responsiveness and are often a barrier to change;
- budgets are rarely strategically focused and often contradictory;
- budgets add little value, especially given the time required to prepare them;
- budgets concentrate on cost reduction and not value creation;
- budgets strengthen vertical command and control;
- budgets do not reflect the emerging network structures that organizations are adopting;
- budgets encourage "gaming" and perverse behaviours;
- budgets are developed and updated too infrequently, usually annually;
- budgets are based on unsupported assumptions and guesswork;
- budgets reinforce departmental barriers rather than encourage knowledge sharing; and
- budgets make people feel undervalued.

The conclusions of the BB group (Hope and Fraser, 2003) are that many of these problems can be mitigated by adopting some new control principles. The first of these principles is designed to remove the emphasis on preset and fixed (budget) targets, and to replace them with benchmarked (flexible) performance standards. Thus, a common form of performance target would

be a league table, either internally (e.g. different branches in the same company) or externally (e.g. performance in comparison with leading competitors) referenced.

This forms the basis for implementing the second, and arguably the more important, principle. Here the objective is to remove reliance on an arbitrary performance target that is set for a fixed period (typically a year) many months in advance. It is to be replaced by a relative performance target that is continually updated in light of changing conditions. More radically, performance against such targets will be evaluated with hindsight. That is, performance evaluations may well be adjusted to reflect the actual operating experience and economic circumstances faced during the period. Rewards may be connected to performance, but, typically, they will be determined by relying more on subjective performance evaluations, with an emphasis on workgroup rather than individual rewards. The aim appears to be to attempt to engender a philosophy of doing what is best for the firm in light of current circumstances, and to encourage teamwork. How successful this approach will prove to be remains to be seen, as it is still largely untested (see Hansen, Otley and van der Stede, 2003, for a review). But it represents a significant change in approach to target setting, in moving away from a preset fixed target to a moving, flexible target benchmarked on other factors. The previous literature on target setting has never considered this possibility, so a new and significant research topic has opened up.

However, the main point made in this section is still valid. That is, the use made of a particular performance measure should determine its operationalization and measurement. Different uses may require (sometimes subtle) differences in definition, and measurement techniques need to be made robust against likely attempts at manipulation. The framework proposed by Otley (1999) and extended by Ferreira and Otley (2005) provides one schema against which any system of performance measures used for management control purposes can be assessed.

Conclusions

Financial and accounting measures of performance often appear to have an objectivity, particularly to unsophisticated users, that turns out to be illusory. The components of any accounting ratio, for example, can be defined in a variety of different ways. No way is objectively

correct or incorrect, but, rather, assessments have to be made concerned appropriateness for a specific use. Even when a ratio has been defined in a conceptually appropriate way there remain issues of measurement. Again, the non-accountant generally has a sense of the objectivity of accounting measurement that is unsupported by its practice. Accounting measures of both cost and profit require a myriad of subjective judgements to be made. For example, the activity-based costing (ABC) literature is replete with examples of the grossly different cost estimates that are produced by traditional and ABC-based cost accounting systems, which may have led to inappropriate product pricing decisions being made. More recently, the EVA[®] literature has proposed well over 100 accounting adjustments that might be made to convert traditional financial accounting numbers, prepared under GAAP, into the most appropriate numbers to be used in the calculation of EVA[®], with EVA[®] to be used as a motivational target for operating managers.

Accounting was once defined, borrowing from a definition of art, as “an attempt to wrest coherence and meaning out of more reality than we ordinarily deal with” (Weick, 1979). Far-fetched as such a comparison may seem, this definition does provide a sense of the complexity of the task being undertaken by much accounting measurement. Financial statements provide, within the confines of a few pages of numerical data, an account of the (financial) outcomes of a complex web of activities undertaken over a period of time. When used for management control purposes the task becomes even more complex, for these accounting measures are intended to help ensure that operating managers are continually motivated and challenged to exercise their managerial skills in the interests of the overall organization. In such a way, the accounting numbers provide a “window” into the organization that gives an (albeit imperfect) image of the activities being undertaken and their consequences.

From such a perspective, the management control function clearly requires an amalgam of both financial and non-financial performance measures, and frameworks for integrating these have been proposed (e.g. the balanced scorecard, the EFQM framework and so on). Within this enterprise, accounting performance measurements should not be treated as a universal “given”, which can be applied in a formulaic manner to any specific situation. Rather, they are like all other performance measures used for this purpose. As such, considerable attention needs to be paid both to their conceptual definition and to the methods of measurement used in their

construction.¹⁰ Thus, accounting performance measures should be neither dismissed nor privileged in the attempt to construct systems of performance management that encourage managers to strive to achieve organizational objectives.

Key research challenges

In an area this broad there are, clearly, many research challenges. But three stand out, and they are reviewed here. First, there is the issue of whether past financial numbers provide a sound basis for predicting future financial numbers, or whether different approaches need to be taken. Second, what are the relative roles of financial and non-financial measures in internal control? Finally, how useful will budgetary control be in the future?

The first issue is very topical. Many people use past financial information as a means of forecasting what the future holds for an organization. At one level, this is entirely sensible; for example, Z-scores based on financial ratios have proven to be a useful tool in predicting bankruptcy. However, at a higher level, this practice is more dubious. It has long been shown that past stock market price information has little or no role to play in forecasting what future stock prices might be. At an intermediate level, the topic presents an interesting research question as market environments become more volatile. What appears likely is that non-financial information from within the firm and information on the wider environment within which the firm operates are likely to provide a better foundation on which to base predictions.

The issue has come up in relation to the marketing of Stern Stewart's EVA[®] performance management system. EVA[®] is an accounting performance measure based on past information; it measures "delivered" performance. Nonetheless, much of the publicity surrounding EVA[®] suggests that it is useful for predicting future performance, or "promised" performance, so-called because future stock prices are based on expectations of what will occur in the future. However, the link between these two concepts is not straightforward. Indeed, the research that has been undertaken suggests that other, more traditional, financial measures are as good predictors of EVA[®] as EVA[®] itself – or better. There is no great reason to suppose that the

¹⁰ Again, the EVA[®] literature provides an excellent case study of how a particular accounting performance measure was adapted and refined from the more basic accounting data found in annual corporate reports.

past will replicate itself in the future, especially as the external environment becomes more uncertain. Thus, modelling the characteristics of an organization in a way that will provide a sound prediction as to its future capability to adapt and prosper in a new environment provides a major topic for research.

Turning to internal control, it is also apparent that traditional financial measures are now seen as an inadequate means of exercising management control within the organization. One of the reasons for the incredible popularity of the balanced scorecard over the past fifteen years has been a realization that non-financial measures provide a much more direct link to the items that managers actually manage. In their more recent publications, Kaplan and Norton have stressed the importance of “strategy mapping”, or constructing the logical connections between the basic non-financial aspects of performance in a chain leading to overall financial outcomes. Research into the validity of such strategy maps is a pressing topic. More fundamentally, most of our financial controls are vertical in orientation – that is, they follow the organizational hierarchy, within vertical “silos”. What has been lacking are horizontal controls that follow business processes or the value chain through the organization. Although we have some non-financial horizontal controls (e.g. cycle time), there is still a great deal of work to be done in devising more such measures.

Finally, despite the problems outlined earlier in this chapter, budgetary control still forms the basis of control activities within most organizations. The debate as to whether the “beyond budgeting” approach can (and should) be applied more widely is still unresolved. In all probability, we need to explore the circumstances under which one approach is likely to be more productive than the other. Thus, whether the answer for a specific organization is to tie budgeting more closely to operational processes, or dispense with it entirely and rely on an array of non-financial measures, is an open and important issue. Indeed, the feasibility of the “adaptive and decentralized” organization advocated by the BB proponents is also questionable. It appears to be applicable to highly divisionalized organizations with little need for internal coordination, but it is not obvious how it should be adapted for use in more centralized organizations with a need for much planning and integration. Again, a contingent answer is likely, but the description of the most important contingent variables remains to be undertaken.

We have learned a great deal about the role and importance of accounting information over the past fifteen years. Much of this learning has been about replacing traditional financial measures with more focused and specific

non-financial measures. But the appropriate balance between the two types of information still requires elucidation, from empirical and conceptual points of view alike.

References

- Bunce, P., Fraser, R., and Woodcock, L. (1995). Advanced budgeting: a journey to advanced management systems. *Management Accounting Research*, 6, 253–65.
- Ferreira, A., and Otley, D.T. (2005). The design and use of management control systems: an extended framework for analysis. Working paper, University of Lancaster.
- Hansen, S.C., Otley, D.T., and van der Stede, W.A. (2003). Recent developments in budgeting: an overview and research perspective. *Journal of Management Accounting Research*, 15, 49–70.
- Hope, J., and Fraser, R. (2003). *Beyond Budgeting: How Managers Can Break Free from the Annual Performance Trap*. Boston: Harvard Business School Press.
- Kaplan, R.S., and Norton, D.P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Lev, B. (1989). On the usefulness of earnings and earnings research: lessons and directions from two decades of empirical research. *Journal of Accounting Research*, Supplement, 153–92.
- Neely, A., Sutcliffe, M.R., and Heyns, H.R. (2001). *Driving Value through Strategic Planning and Budgeting*. London: Accenture.
- Otley, D.T. (1999). Performance management: a framework for management control systems research. *Management Accounting Research*, 10, 363–82.
- Weick, K.E. (1979). *The Social Psychology of Organizing*, 2nd edn. London: Addison-Wesley.
- Young, S.D., and O’Byrne, S.F. (2001). *EVA[®] and Value-Based Management: A Practical Guide to Implementation*. New York: McGraw-Hill.

2

Measuring marketing performance: research, practice and challenges

Bruce Clark

Introduction: the pressure to measure

It is a heady time for marketing performance measurement. In the past several years the pressure to measure the performance of marketing better has become relentless from both corporate managers and the academics and consultants who work with them. In the United Kingdom several marketing trade organizations recently came together to sponsor the Marketing Metrics project (Ambler, 2003), and the Chartered Institute of Marketing issued its own substantial report on marketing effectiveness (Chartered Institute of Marketing, 2001). In the United States the influential Marketing Science Institute declared marketing metrics a key research priority in four consecutive biennial reports (Marketing Science Institute, 1998, 2000, 2002, 2004), supported by several well-attended joint conferences of academics and practitioners (Debruyne and Hubbard, 2000; Bahadir and Tuli, 2002; Bruno, Parthasarathi and Singh, 2005) and a special section of the October 2004 issue of the *Journal of Marketing* (Lehmann, 2004). The American Marketing Association followed up with a major study among its members on marketing accountability (American Marketing Association, 2005). Other US-based organizations, such as the Marketing Leadership Council (Marketing Leadership Council, 2001, 2002, 2003) and the CMO (chief marketing officer) Council (CMO Council, 2004), have also issued their own reports on improving performance measurement.

This rush of interest has been driven by a number of factors. Perhaps first and foremost, the post-2000 economic downturn sent many corporate leaders on a mad scramble to cut costs, and marketing budgets are notoriously seen as soft targets in this regard (Sheth and Sisodia, 1995). One estimate suggests that marketing expenses represent between 5 per cent and

over 20 per cent of corporate revenues on a given income statement (Eechambadi, 2005): a target that is fat as well as soft. Furthermore, the marketing budget that has remained is now subject to greater scrutiny; the “tyranny” of the quarterly profit and loss account is a real issue for chief marketing officers at any publicly held organization (Webster, Malter and Ganesan, 2005).

On a more optimistic note, attention to multidimensional performance measurement schemes (e.g. the balanced scorecard: Kaplan and Norton, 1992, 1996) has raised interest in measuring the customer perspective on firm performance, typically related to marketing. Similarly, technological solutions in the area of customer relationship management have made it possible for more firms to assess the effects of their marketing activities (Kim, Suh and Hwang, 2003). Investors and analysts are also increasingly asking for information on the marketing performance of firms (Haigh, 1998; Mavrincac and Siesfeld, 1997); accounting rules regarding the capitalization of brands, in particular, have recently become much more specific in regard to reporting (Haigh and Knowles, 2004).

This chapter focuses on reviewing recent developments and approaches in measuring marketing performance: research and practice have evolved dramatically over the last few years. A more historically oriented review is available in the previous edition of this book (Clark, 2002) or in Bonoma and Clark (1988).

Defining the problem: what is marketing?

Assessing marketing performance is very difficult. Unlike purely internal measures of performance, such as defects per million, marketing performance depends on external, largely uncontrollable actors, such as customers and competitors. Moreover, it acts as a mediator between these external actors and various internal corporate processes, such as accounting, operations, research and development, and finance. Bonoma and Clark (1988, 2) observe that these factors make “marketing’s outputs lagged, multivocal, and subject to so many influences that establishing causes-and-effect linkages is difficult”.

At a basic level, part of this arises from the definition of marketing itself. One cannot measure what one cannot define. Two definitions from well-respected organizations are as follows.

- (1) Marketing is an organizational function and a set of processes for creating, communicating and delivering value to customers and for managing

customer relationships in ways that benefit the organization and its stakeholders (Keefe, 2004).

- (2) Marketing is the management process responsible for identifying, anticipating and satisfying customer requirements profitably (Chartered Institute of Marketing, 2005).

These definitions highlight a number of the issues that make marketing so problematic to measure. First, there is the distinction between function and process. Piercy (2002) eloquently makes the case that too much attention has been paid to the status of marketing departments within organizations, and too little to the processes by which organizations go to market. The budget arguments inside organizations, on the other hand, are very much oriented to the productivity of the marketing department's budget (Webster, Malter and Ganesan, 2005). Second, there is the sheer difficulty of measuring the processes. How do we measure customer value? How do we measure the quality of a customer relationship? How do we measure the quality of the processes that lead to improvements in these two judgements? Finally, there is the link to profit. While work on customer profitability and marketing ROI has advanced substantially in the past several years, it is still difficult to do well and can lead to perverse incentives (Ambler, 2003; Webster, Malter and Ganesan, 2005).

We begin by reviewing the traditional approach to measurement: marketing productivity.

The traditional approach: marketing productivity

Historical approaches

From the earliest studies through to the 1970s, the vast bulk of the work on measuring the performance of marketing looked at marketing productivity. Drawing on earlier work in economics and manufacturing productivity, these efforts typically looked at measuring output per unit of input as a means of assessing marketing's contribution to the success of firms and industries. Work in this area explicitly attempted to integrate finance and accounting perspectives, especially cost accounting, into evaluating the marketing function. A typical approach involved comparing the productivity of different marketing programmes within a firm to inform resource allocation decisions across programmes. Programmes might represent sets of marketing activities performed to sell particular products or serve particular

market segments. Inputs would represent either marketing expenditures or marketing activities (e.g. advertising or sales force efforts) devoted to those programmes.

The historical roots of this work in marketing date back to the 1960s (see Bonoma and Clark, 1988, for a detailed review). Sevin's *Marketing Productivity Analysis* (1965) is a minor masterpiece of the marketing productivity literature, spending a little over 100 pages to lay out detailed profitability analysis for products and marketing programmes. Feder (1965) borrowed from the microeconomic literature to discuss comparing marginal revenues to marginal costs as a way of better allocating marketing resources. Goodman (1970, 1972) followed in Sevin's footsteps by examining profitability and the return on investment of marketing activities, but made perhaps his most intriguing contribution to the literature in his advocacy of establishing the position of "marketing controller" within firms (1972). Later work in this stream looked at more sophisticated financial approaches to evaluating marketing productivity. Influential articles by Buzzell and Chussil (1985) and Day and Fahey (1988) pioneered the use of cash flow to measure marketing performance and its contribution to shareholder value.

Productivity research continued into the 1990s, typically with more elegant analytic techniques as an aid. Data envelopment analysis has attracted particular attention as a way of mapping different organizations' or business units' efficiency at reaching a multidimensional performance frontier (e.g. Bhargava, Dubelaar and Ramaswami, 1994). Sheth and Sisodia (1995) combined several approaches to suggest that true marketing productivity should reflect the amount of *desirable* output per unit of input.

Marketing return on investment

Marketing return on investment (ROI; or return on marketing investment, ROMI) is the latest heir to productivity research. It harks back to the pure input-output model, and has the virtue that it denominates marketing performance in a way that the financial officials of an organization understand: they dispense budget and expect returns. In light of the pressure to measure indicated earlier, what could be more natural than measuring marketing performance as return on investment? Managerial writing (and consulting) on the subject is voluminous (e.g. Lenskold, 2004; Neff, 2005), and various organizations, ranging from Kraft (Duffy, 2000, 2002) to Colgate (Bacon, 2002) to Fidelity Investments (Kincaid, 2000), claim to be gaining benefits from measuring ROI. Scholars have also weighed in with

more nuanced models that nonetheless retain the ROI logic (e.g. Rust, Lemon and Zeithaml, 2004).

Despite this popularity, there are some drawbacks to ROI (see Ambler, 2003, for a particularly forceful discussion). First, the term itself has become interchangeable with “accountability” in many organizations (Neff, 2005), meaning that it can be measured differently despite a consistent label. Second, it tends to be measured on a short-term basis (e.g. this year’s return divided by this year’s investment), which is problematic in marketing expenditures from which we expect long-term effects. Finally, and perhaps most seriously, using ROI as a performance measure encourages underspending (Ambler, 2003; Kumar and Petersen, 2004). Economics tells us that the way to maximize profit is to continue an activity so long as its marginal revenues exceed its marginal costs. As diminishing returns set in on any given marketing tactic, revenues will exceed costs by less and less: total profit will continue to rise, but the ratio of returns to spending will gradually drop. ROI is particularly problematic if managers receive compensation based on it. The temptation to cut corners on investment in the current period while hoping that returns will persist until the next period is high.

A chain of effects

While input-output models made for a plausible first cut at performance measurement, it is widely acknowledged that a more sophisticated view implicates marketing in some kind of “chain of effects”. In this approach, marketing activities lead to some kind of customer reaction, either attitudinal or behavioural, which in turn leads to sales and profit, ultimately affecting shareholder wealth (Ambler, 2003; Lehmann, 2004; Rust *et al.*, 2004). Much research in marketing has looked at the impact of marketing and other activities on the intermediate customer reaction variables – e.g. customer satisfaction (Szymanski and Henard, 2001) or attitude to the brand (Keller, 2002) or purchase intention/purchase (Mittal and Kamakura, 2001). It is only more recently that firmer links to financial or stock performance have been modelled (e.g. Aaker and Jacobson, 1994, relating brand attitude to stock prices; Anderson, Fornell and Mazvancheryl, 2004, and Gruca and Rego, 2005, relating customer satisfaction to shareholder value).

Beyond a chain of effects, it is clear that performance measurement provides feedback to the organization, influencing both management behaviour and capabilities (Clark, 1999; Greve, 1998; Lant, 1992; March and Sutton, 1997; Miller, 1994; Rust *et al.*, 2004). Combining these perspectives allows

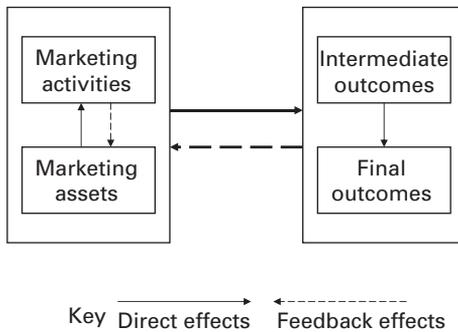


Figure 2.1: A framework for marketing measurement

us to provide an organizing framework around marketing measures, as indicated in figure 2.1.

The remainder of this chapter uses figure 2.1 to look, first, at what to measure in each of the indicated areas; there is voluminous research and practitioner advice in this regard. We shall then move to a broader discussion of issues and challenges to marketing measurement research and practice.

What to measure?

Marketing activities

Marketing activities for the organization have traditionally been organized around the “marketing mix”, consisting of product (creating value for the customer), pricing (charging appropriately for that customer), promotion (communicating value to the customer) and place (delivering value to the customer). More recently separate measures for customer service have become common, as post-sale service has become more important in many industries.

Measurement of activities is to some extent a documentation exercise: how much did we spend on key activities, and have we executed them in the time frame we planned? These activities are then related to intermediate and outcome measures, often in the productivity or ROI paradigm previously discussed. Marketing mix modelling is increasingly popular as a means of conducting these analyses (Doyle, 2004; Neff, 2005), although it is methodologically complicated (as shown by Manchanda, Rossi and Chintagunta, 2004) and requires large amounts of data. It is used to evaluate past performance, adjust resource allocation across activities and predict future performance.

Table 2.1: Representative marketing activities and measures

Activity	Measure
Product	Number of new products introduced last year; number of stock-keeping units in portfolio
Price	Price relative to key competitors; markup on cost; retailer margin
Promotion	Number of promotional events held; number of sales calls made; number of advertising exposures
Place	Number of distribution outlets; number of products carried per outlet
Service	Delivery time; percentage of perfect orders (correct products on time); mean time to resolve customer problem

Aside from expenditure, activities vary by element of the mix, organization and industry. A representative set of activities is presented in table 2.1. Generally, managers and academics are interested only in these activities in terms of how they affect outcomes, but they must be measured accurately nonetheless. Particular progress has been made in measuring sales force activities in the context of sales force automation and customer relationship management software, though the usual teething problems with software implementations apply (e.g. Honeycutt, 2005).

Intermediate outcomes

Before customers purchase a product (a) the organization has to create, communicate and make available the product, and (b) customers have to believe that the product will provide value to them relative to competing offerings. These intermediate outcomes have been the subject of much research and practice in marketing, usually within the context of a specific marketing tool.

At a basic level, customers have to be aware that a product exists. Awareness is a common measure of both advertising effectiveness and more general marketing programme success (Ambler, 2003; Marketing Leadership Council, 2001; Winer, 2000). It is usually measured as either unaided (“List all the laundry detergents you can think of”) or aided (“Tick all the laundry detergents in this list that you recognize”) awareness.

Drilling down from basic awareness, many firms move on to measure what customers know and feel about the product. Familiarity with the product is a common measure, and more detailed readings regarding knowledge about specific product attributes or benefits can be taken. It is also increasingly

common to measure associations in memory (“When you think of Ariel, what is the first word that comes to mind?”) and emotions regarding the product.

Knowledge and emotions translate into customer attitudes regarding the product: consideration (“Would you consider purchasing Ariel the next time you purchase laundry detergent?”), liking (“On the following scale, indicate how positively or negatively you feel about Ariel”) and preference (“My favourite detergent brand is ____”) are common intermediate measures in this regard.

The above sequence is sometimes called a “hierarchy of effects” in communication research. While it suggests a law-like progression from cognition to affect (and purchase), in reality customers do not follow a uniform sequence, and alternative sequences may apply to different customers in different product categories. However, it remains a useful conceptual framework for organizing measurement regarding intermediate outcomes (see Weilbacher, 2001, and Barry, 2002, for a discussion).

Specific research regarding price and availability are useful measures of intermediate performance. Perceived price relative to value received (“Ariel is good value for what I pay”) and relative to competing options (“How much more or less would you be willing to pay for Ariel when compared to Sainsbury brand detergent?”) are good early indicators of whether pricing is appropriate for the product.

In terms of availability, organizations want to know the degree to which their product is available in the right locations at the right time for customers. Increasing attention has been devoted to this issue in the context of more sophisticated supply chain management (e.g. Gunasekaran, Patel and McGaughey, 2004). Measures here include penetration of target distribution channels (% of outlets in a channel through which products are carried), shelf space, stockouts and delivery time (Winer, 2000). In addition, one can consider applying awareness, cognition and affect measures to distribution channel members.

Final outcomes

After all the creation, communication and delivery of value to the right place at the right time, did the customer buy? Sales (turnover), either in value or volume, have historically been a dominant outcome by which most marketing programmes (and marketers) have been judged (Bonoma and Clark, 1988). Sales routinely continues to appear at or near the top of any list of

marketing metrics that firms follow (e.g. Ambler and Riley, 2000; Clark, 2000; Marketing Leadership Council, 2001; Winer, 2000).

Benchmarking sales against competitors in the form of market share is also enormously popular (Ambler and Riley, 2000; Clark, 2000; Winer, 2000). From the late 1970s through to the late 1980s unit market share attracted tremendous attention as an output variable. Work by the Boston Consulting Group (Henderson, 1973) and the Profit Impact of Market Strategies (PIMS) project (Buzzell and Gale, 1987) concluded that market share was a strong predictor of cash flow and profitability. In retrospect, the relationship between market share and profitability has proved more complicated (Jacobson, 1988; Szymanski, Bharadwaj and Varadarajan, 1993). There is also some evidence that the competitive focus that market share measures engender can be counterproductive to profitable decision making (Armstrong and Collopy, 1996).

The problem with taking sales as a performance measure is that the same result can be achieved by two means: better products or lower prices. The latter frightens chief financial officers (CFOs), who already suspect that marketing is more concerned with acquiring customers than making money. To be fair, salespeople in particular have usually been compensated based on their sales; one can hardly complain about the focus on making the sale that results.

Nonetheless, profitability is now equal with sales as a performance measure on many lists, especially among measures reported to the top executive board (Ambler and Riley, 2000). Pressure to produce sales and customers in a profitable manner is a large portion of the demand for accountability in marketing. 84 per cent of respondents to the American Marketing Association's accountability study indicated that it was extremely or very important to link marketing programmes to financial performance, and that pressure to demonstrate accountability was higher for senior marketing executives in the organization (American Marketing Association, 2005). Whether in terms of total profit, profit margins or profit ratios (i.e. ROI, return on assets (ROA) or return on sales (ROS)), profit now features on marketing executives' agendas at a level previously unseen.

Marketing scholars have increasingly been attempting to document the links between marketing activities and shareholder wealth. Three possible measures have been discussed: cash flow, the market-to-book value of the firm and stock price reactions.

Srivastava, Shervani and Fahey (1999) lay out an important theoretical rationale for cash flow as a measure of marketing performance, claiming that marketing can have four positive influences on cash flow. First, of course,

marketing can increase the amount of cash flowing into the organization through increased sales. Second, marketing can shorten the time it takes for cash to enter the organization by using promotions to accelerate purchases. Third, marketing can reduce the volatility of cash flows, thereby reducing financial risk, typically through a contractual purchase scheme. Finally, marketing investments in assets such as a brand can increase the long-term value of the business; Ambler (2003, 5) notes that a brand is essentially an “upstream reservoir of cash flow, earned but not yet released to revenue”.

Other scholars have looked at market capitalization or stock market measures of performance as they relate to marketing. Tobin’s *q*, which looks at the market capitalization of the firm relative to the replacement value of total assets, has been a popular measure in this regard, used as a dependent variable in studies of branding (Simon and Sullivan, 1993), customer satisfaction (Anderson, Fornell and Mazvancheryl, 2004) and technological innovation (Lee and Grewal, 2004).

Finally, stock market reactions to marketing events have been used to make marketing–finance links explicit. Typical event studies will look at “excess stock market return” – i.e. an abnormally high stock price – to a particular event over a short time period. Lane and Jacobson (1995), for example, have found both positive and negative stock price reactions to brand extension announcements depending on brand attitude and familiarity among consumers. Other topics investigated in this fashion include reactions to announcements regarding new internet channels (Geyskens, Gielens and Dekimpe, 2002), Olympic sponsorships (Miyazaki, 2001), new product development announcements (Sharma and Lacey, 2004), American Super Bowl advertisements (Kim and Morris, 2003) and unanticipated shifts in marketing strategy for the firm (Mizik and Jacobson, 2003).

Marketing assets

There has been recent and continuing attention paid to the notion of developing good marketing assets (Ambler, 2003; Piercy, 1986; Srivastava, Shervani and Fahey, 1998). Piercy defines an asset as a “value-producing resource” for the firm. Srivastava, Shervani and Fahey suggest that assets can be divided into relational and intellectual assets, the former covering relationships with current external stakeholders (e.g. customers, channels) and the latter covering the knowledge that the firm has about its environment. These may accrue to the company as a whole or to specific business units. The most valuable assets typically take time to develop; if inimitable,

they can represent a significant advantage in the marketplace. An asset-based perspective on marketing suggests that good marketing develops good marketing assets, which in turn can be leveraged to generate superior business performance over the long term. Primary research in this area has examined two kinds of assets: brand-based and customer-based.

The brand asset

Many researchers and managers believe that a powerful brand is among the most important marketing assets a firm can manage (see Barwise, 1993, and Keller, 2002, for reviews). Strong brands, it is argued, (1) allow firms to charge price premiums over unbranded or poorly branded products, (2) can be used to extend the company's business into other product categories (e.g. the Ivory brand name, originally used on soap, was extended to introduce Ivory Shampoo) and (3) reduce perceived risk to customers (and, perhaps, investors). All this should produce higher returns for the firm owning the brand.

The strength of a brand represents its "equity" in the marketplace. Measuring this strength has typically taken two different approaches. The behavioural, or customer-based, approach looks at customer response to the brand, either in terms of perceptions or purchase. A representative definition of behaviourally based brand equity is the differential effect of brand knowledge on customer response to marketing of the brand (Keller, 1993). Customers in behavioural studies typically respond more favourably to strong brands than to unbranded or poorly branded products. Measurement in this area consists primarily of survey questions tapping underlying attitudes towards the brand regarding issues such as quality and value (e.g. Yoo and Donthu, 2001; Vazquez, Del Rio and Iglesias, 2002; Netemeyer *et al.*, 2004). Many of the hierarchy of effects questions discussed previously fit easily into this framework. Consulting firms and advertising agencies have busily developed their own behavioural measures with which they can solicit and inform clients (e.g. Millward Brown's BrandZ, Young and Rubicam's Brand Asset Valuator). Alternative approaches attempt to look at purchasing behaviour or willingness to pay a revenue premium for a brand (see Ailawadi, Lehmann and Neslin, 2003, for an example of the latter).

The financial approach to brand equity attempts to divine the financial value of the brand to firms and their investors. A widely cited approach in this area has been developed by Simon and Sullivan (1993), who define brand equity as the incremental cash flows that accrue to branded products over and above the cash flows that would result from the sale of unbranded

products. They use Tobin's q to build a valuation measure. Higher cash flows should in turn lead to higher stock market valuations (e.g. Aaker and Jacobson, 1994; Kerin and Sethuraman, 1998).

Measuring brands through their impact on the stock market is a growth industry, aided by changes in accounting standards that have forced corporations to value acquired brands separately on their balance sheets (Haigh and Knowles, 2004). Interbrand's valuation and rankings (www.interbrand.com) are perhaps the most widely publicized financial method in the practitioner realm. They use an earnings multiple calculated through a proprietary methodology to value brands.

There is little question that brands can make a powerful difference in how customers respond to brands and brand extensions, but defining and measuring the brand continues to elude consistency. Ambler (2003) calls the brand an "elephant", recalling the story of the blind men who attempted to divine the core nature of an elephant while feeling different parts. DeChernatony, Riley and Harris (1998) interviewed twenty brand experts and found that they named between one and eight measures of brand success each. Ambler and Barwise (1998) also note the frequent confusion between the definition of brand equity and its measurement (e.g. does the financial value of a brand represent its definition or a measurement of its strength?). Because brand equity usually accumulates and dissipates over long periods of time, it is also difficult to use it as a short-term performance measure.

The customer asset

The other marketing asset that has attracted extensive attention has been the customer base. Just as a firm can build or acquire brands that have long-term value, the firm can also build or acquire a customer base, on which it hopes to rely for repeat sales, purchases of new products, and referrals to prospective customers.

Customer satisfaction

The move to regarding the customer base as a key marketing asset has built up in layers. First it was driven by customer satisfaction research and practice, in the late 1980s and early 1990s (see Halstead, Hartman and Schmidt, 1994, and Yi, 1990, for reviews). Through the 1990s customer satisfaction became an important benchmark measure for many firms

(Rigby, 2001); it featured prominently as a plausible measure to include in a balanced scorecard (Kaplan and Norton, 1996).

The basic notion behind customer satisfaction is that customers have expectations about the products and services they buy, and are more or less satisfied depending on how well the consumption experience meets or exceeds those expectations – sometimes called the “disconfirmation of expectations” paradigm. Having a satisfied customer base is considered an important marketing asset, because it should lead to increased loyalty, with its consequent revenue implications (see below) and lower marketing costs. As with brand equity, a satisfied customer base is presumed to be a leading indicator of future success. Measurement of satisfaction is typically accomplished by surveys of the customer base.

Research on satisfaction as both measure of success and predictor of success has proved to be maddeningly uneven in outcome. Tests of the disconfirmation of expectations paradigm have produced mixed results, leading to multiple competing satisfaction frameworks (e.g. Anderson and Sullivan, 1993; Teas, 1993; Voss, Parasuraman and Grewal, 1998). Controversy has arisen around the correct measurement of expectations (Teas and Palan, 1997), and whether one must measure multiple aspects of satisfaction with a product, either in terms of multiple processes (Spreng, MacKenzie and Olshavsky, 1996) or individual product attributes (Donaher and Mattson, 1994; Halstead, Hartman and Schmidt, 1994).

That said, the American Customer Satisfaction Index, administered by the University of Michigan, and its counterparts in other countries have become highly influential. Its key survey items measure overall satisfaction, disconfirmation of expectations, and performance relative to ideal for the product category (Fornell *et al.*, 1996). Estimates of these variables for firms have been positively associated with profitability, cash flow and shareholder value (Anderson, Fornell and Mazvancheryl, 2004; Gruca and Rego, 2005), though there appears to be substantial variation across industries. More broadly, a meta-analysis by Szymanski and Henard (2001) concludes that satisfaction has many positive effects.

Customer loyalty

Nevertheless, a challenge arose in the 1990s from individuals arguing that customer loyalty was far more important than customer satisfaction as a leading indicator of firm performance. Advocates of loyalty observed that it is not whether customers are satisfied that affects cash flow, it is whether they

remain customers of the firm over time. Reichheld (1994) suggests that good marketing attracts the right customers: ones whose loyalty the firm is able to earn and keep.

A loyal customer base should be an important marketing asset for several reasons (Dick and Basu, 1994). Loyal customers are easier to retain, so marketing costs for these customers should be lower; they are less likely to search out information on competing products; and they are more resistant to persuasion efforts by competitors. Given retention, over time firms hope to obtain more business per loyal customer, as the customer's favourable initial impression of the firm leads to a willingness to try other products. Loyal customers may be willing to pay a price premium. Finally, having loyal current customers may reduce the acquisition cost for new customers through positive word-of-mouth effects, although the impact of these may decline with the duration of the relationship (see East, Lomax and Narain, 2001).

Loyalty is defined and measured in two ways. One is attitudinal loyalty, indicating the degree to which a customer desires to be loyal to a firm. This is typically established using survey measures of attitude and repurchase intention. The other is behavioural loyalty, indicating the degree to which the customer is actually loyal to the firm. This requires tracking purchases through a customer database. A popular model to segment a customer base by loyalty is to look at the recency, frequency and monetary amount of purchases each customer makes with the organization (the "RFM" model: Verhoef *et al.*, 2003). The duration of the relationship, usage rates and cross-buying are also common behavioural indicators of loyalty (Bolton, Lemon and Verhoef, 2004). The higher the customer is on each of these measures, the more he or she is deemed loyal and valuable.

For contractually based goods or services, such as mobile phones, the financial benefits of loyalty to the firm are obvious: every incremental month is an incremental tick in revenue. Research in non-contractual settings, however, is less positive about the likely benefits of loyalty. Reinartz and Kumar (2000) find no evidence that long-lived customers in these settings are more profitable than short-lived ones. As with satisfaction, it appears that the benefits to the firm of having loyal customers vary by product category.

Customer relationship management

Satisfaction and loyalty research have come together in the burgeoning area of customer relationship management (CRM). An heir to database marketing, CRM uses analytic techniques to segment customer bases extremely

finely, and then customizes offers to each segment. Behavioural variables of the kind discussed under customer loyalty are the starting place for this kind of approach, which relies on sophisticated software applied to large databases. UK retailer Tesco, for example, segments its Clubcard database using as many as forty dimensions to identify each customer's "DNA profile", and then sends more than four million variants of its Clubcard statement to its ten million customers every quarter (Humby, 2005).

In terms of measuring the value of this customer database, the dominant metric has been customer lifetime value (CLV: Gupta and Lehmann, 2003; Gupta, Lehmann and Stuart, 2004; Hogan *et al.*, 2002). The lifetime value of an individual customer is simply the sum of the discounted value of the cash flows generated by the customer,

$$CLV = \sum_{t=1}^n \frac{m^t}{(1+i)^t}$$

where m is the margin in a given time period t , n is the number of periods over which the customer remains active and i is the discount rate. The worth of a customer base is then the sum of the individual CLVs. While simple in concept, actual estimation of this metric requires knowledge or estimates of both per period margins and the length of time the customer keeps purchasing from the organization. The retention rate for a firm – the proportion of customers who keep purchasing in a given time period – is notoriously hard to estimate, leading to a variety of short cuts and simplifying assumptions being used in research (Gupta and Lehmann, 2003).

Nonetheless, a large software industry and research practice have developed around this measure to help firms measure performance, predict it and allocate resources. Measurement of this kind is clearly useful, but it is also difficult to do, especially for small firms or large firms in new businesses with which they have little customer history. This relates to the general point that loyalty research, to some extent, has been better at describing what to do once one has a loyal customer base than it has at describing how to obtain such a base.

Measurement issues and challenges

Given a cycle of assets, activities, intermediate outcomes and final outcomes, several practical and theoretical issues arise as we move forward.

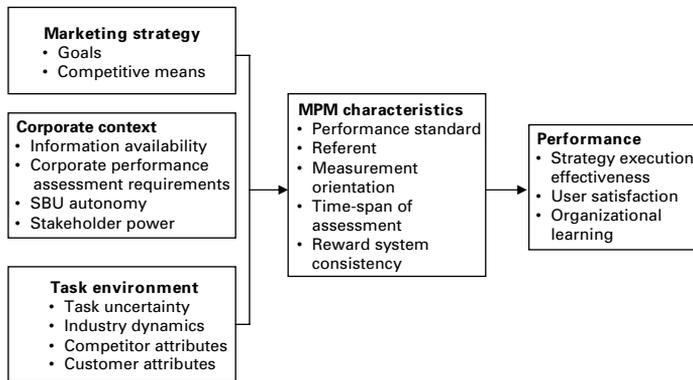


Figure 2.2: A contextual framework for marketing performance measurement (MPM) systems

Source: Adapted from Morgan, Clark and Gooner (2002).

Developing effective marketing performance measurement systems

Morgan, Clark and Gooner (2002) lay out a theoretical framework for considering how to build effective marketing performance measurement (MPM) systems. They point out that a gap exists between what academics might generally provide as a “normative” model for performance measurement and the “contextual” situation in which individual businesses find themselves. Effective performance measurement is a contingent process: the nature of a performance measurement system will change depending on the firm’s marketing strategy, corporate context and task environment. Their model is presented in figure 2.2, and suggests the raft of considerations any organization must address to be successful.

The corporate context that Morgan, Clark and Gooner cite seems particularly likely to be important. In terms of information availability, some organizations are overwhelmed with data, while others go begging. Performance assessment requirements also vary dramatically across firms depending on the attitude of the top management board, industry norms and regulatory regimes. This relates to the two remaining issues that Morgan, Clark and Gooner (2002) raise. Strategic business unit (SBU) autonomy indicates the degree to which a particular business unit may design a system independent of the corporate centre. Stakeholder power reflects the fact that many different groups beyond executives have an interest in the nature of performance measurement and reporting: employees, regulators, customers and marketing partners (e.g. suppliers, advertising agencies) all have an interest in the nature of marketing activities being performed by an organization. One recent trend

is increased scrutiny by non-governmental organizations regarding the supplier and marketing practices of organizations. The fair trade movement is a good example of the former, while the latter includes the increasingly strident criticism of the advertising and distribution of “unhealthy” products in the United States (soda, alcohol, tobacco). That stakeholders may have incompatible interests only complicates the picture of which organizations perform “the best”.

While advice abounds regarding the question of how to measure effectively, relatively little rigorous empirical research has been conducted to examine the implementation of MPM systems. Clark, Abela and Ambler (2004) have found that developing good measurement practices in the areas of brand equity and the measurement of financial returns are positively related to both firm performance and learning. Some recent research suggests that marketing dashboards have positive benefits for the dissemination and use of MPM system information (Abela, Clark and Ambler, 2005; Reibstein *et al.*, 2005). Reinartz, Krafft and Hoyer (2004) have found that effectively aligning the organization around customer-related processes has more powerful positive benefits for CRM programmes than the extent to which CRM technology has been adopted. Beyond anecdotes, Morgan, Anderson and Mittal (2005) break the customer satisfaction measurement process into data scanning, analysis, dissemination and utilization, and find qualitative evidence that superior process relates to superior performance. They suggest a large variety of contingencies associated with the implementation of measurement processes, grouped under organizational culture, customer and competitor power, resources and strategy.

More research pending, the fundamental points here seem to indicate that effective MPM system design is a contingent process, and that there must be an alignment between business processes and measurement for the latter (and perhaps the former) to succeed.

Understanding multiple purposes of measurement

Related to implementation is the purpose of measurement. Scholars have applied little systematic discussion to the purposes to which marketing executives might put measurement. Reading the literature, one might be forgiven for thinking that the sole purpose of marketing measurement is to demonstrate that marketing activities are profitable, reflecting the intense pressure marketers find themselves under to justify their worth

(e.g. Ambler, 2003; Webster, Malter and Ganesan, 2005). Recalling our discussion of the definition of marketing and marketing ROI, the purpose of much marketing measurement seems to be heavily focused on evaluating what Ambler (2003) calls “budgetary” marketing, the return on marketing dollars spent.

Ailawadi, Lehmann and Neslin (2003, 2) suggest five general purposes in measuring brand equity: (1) to guide decisions, (2) to assess the extendability of the brand, (3) to evaluate the effectiveness of marketing decisions, (4) to track the brand’s health, and (5) to assign a financial value to the brand. With the exception of (2), most of these would apply to overall marketing measurement in general.

More broadly, Menon and Varadarajan (1992) suggest that market information use can be divided into instrumental, conceptual and symbolic use. Narrowing this to marketing performance information, instrumental use involves the direct application of performance information to solve a particular problem (e.g. “Should we increase advertising for our new product?”). Conceptual use may be thought of as more general development of a managerial knowledge base (e.g. “What is the general relationship between advertising and new product success in our firm?”). This is the classic organizational learning benefit of measurement (e.g. Morgan, Clark and Gooner, 2002). We measure to develop our ability to learn what works and therefore to conduct marketing more effectively in the future. Marketing mix models have long been used for this purpose; more recently, customer value models have been proposed as another way to guide future resource allocation (Venkatesan and Kumar, 2004). Finally, symbolic use occurs when managers distort marketing information for their own purposes (e.g. “Which performance measures support my desire to increase advertising?”).

Marketing could profitably learn from work in other areas regarding a systematic treatment of measurement purpose. As a general framework, Neely (1998) lays out four purposes of measurement: (1) to check an organization’s position, (2) to communicate that position to important stakeholders, (3) to confirm priorities for decision makers, and (4) to compel progress by motivating effort. Other frameworks and sets of purposes for measurement may be found from the perspectives of accounting (Kaplan and Norton, 1992, 1996; Shank, 1989), public administration (Behn, 2003), research and development (R&D) management (Kerssens-van Drongelen, Nixon and Pearson, 2000) and human resource management (Toulson and Dewe, 2004).

Understanding feedback loops

As we use MPM systems more and more for prediction and resource allocation, feedback from performance to activities becomes more important (cf. Dickson, Farris and Verbeke, 2001). Not only do activities create assets and outcomes, they are also created by them. Indeed, the point of creating a marketing asset is to exploit it.

Psychologically, previous success or failure can have profound consequences for further managerial behaviour (Miller, 1994). Strategically, an important referent for managers is their goal or aspiration. To the extent that they succeed or fail to meet their aspirations, both management behaviours and aspirations adapt (e.g. Greve, 1998; Lant, 1992).

Econometrically, feedback loops raise the question of whether marketing actions can truly be modelled as exogenous of performance (Manchanda, Rossi and Chintagunta, 2004). Standard regression models, for example, assume that performance is affected by activities, but not that activities are affected by performance. To the extent that this is unlikely to be true, different techniques should be used, especially with time series data. Two-stage least squares is a standard technique in this regard (Greene, 1990), but a recent marketing literature has also emerged around vector autoregressive models (see Franses, 2005, for a discussion).

Reconciling multiple dimensions of performance

A continuing struggle in performance measurement generally and marketing performance in particular is the reconciliation of the many different measures of performance available. That performance is multidimensional is well established (Lewin and Minton, 1986; Bhargava, Dubelaar and Ramaswami, 1994) but it can make managers' lives difficult, either through inconsistency among metrics or through simple information overload.

As noted previously, managerial use of dashboards and stoplight charts appears to be on the increase, thanks in part to software availability. Theoretically, putting multiple measures into a chain of effects model specifies relationships, but does not necessarily make the "what" and "how" to measure questions clearer (Morgan, Clark and Gooner, 2002). Statistically, multivariate techniques such as multivariate analysis of variance/multivariate regression, data envelopment analysis and structural equation modelling can handle multiple dependent variables, and factor analysis can be an effective data reduction technique among correlated measures, but

these typically represent too technical a hurdle for widespread managerial use. Specifying more clearly the relations among performance variables would help improve the situation (e.g. Selnes, 1993; Clark, 1999).

Reconciling short- and long-term measurement

Perhaps the most frustrating problem of multiple performance measures for marketing executives is the perceived conflict between short-term activities, short-term performance and long-term performance. Marketers commonly feel that the long-term investment aspect of marketing spending (e.g. in customer or brand terms) is undervalued by the rest of the firm (Webster, Malter and Ganesan, 2005). Plaintive calls that accounting rules should be changed to capitalize at least some marketing expenditures reflect this managerial conviction.

While it is clear that the impact of some marketing activities changes depending on the time over which the effect is measured, evidence on the balance of long-term and short-term impacts from marketing has been mixed (e.g. Hanssens, 2005). Perhaps the best practical suggestion in this regard comes from Ambler (2003), who argues that marketing performance should consist of a measure of short-term performance adjusted for change in the value of long-term marketing assets.

Summary

Returning to the framework presented in figure 2.1, we can summarize several sets of conclusions regarding marketing measurement practice and challenges.

Marketing activities

The measurement of marketing activities is to some extent a documentation exercise that, nonetheless, provides the input to understanding the effects of marketing: these are the marketing “business processes” to which we must align our measurement. IT packages increasingly automate this exercise, but the critical assumption is that the correct activities are being measured in the correct fashion. The “4 Ps” marketing mix framework (product, price, promotion, place) is a common one for categorizing activities; other activities (especially service) may apply, depending on the situation.

“Success” or “strategy” maps may be useful as a tool to identify key activities for measurement.

Intermediate outcomes

The intermediate outcomes to be measured comprise the cognitive and behavioural reactions that customers exhibit to marketing activities prior to purchase. The key objective here is to identify those intermediate outcomes that are leading indicators of purchase and retention. Knowledge, attitude and pre-purchase behaviours (e.g. store visits) are all likely measure categories to investigate in this respect.

Final outcomes

The dominant measures of marketing performance remain final outcomes such as sales and profitability; if marketing executives measure nothing else, they measure one or both of these, either in an absolute (units, value) or ratio (market share, marketing ROI) sense. While ratio measures such as ROI can present problems, generally it is highly appropriate to attempt measurement of the final outcomes that arise from marketing efforts. Beyond the income statement, scholars are increasingly attempting to link marketing activities to shareholder wealth, but, with the exception of brand valuation, this has not yet significantly penetrated management practice.

Marketing assets

The long-term value-producing resources for the firm are usually conceptualized in terms of brand or customer assets, but one may think of other knowledge or process capabilities that have similar long-term benefits. Both brand and customer measurement have received extensive attention in research and practice over the past decade. In terms of brands, behavioural and financial metrics have been developed. A challenge in this area is reconciling these differing conceptions to a consistent definition of the brand asset. Regarding customers, measures and linkages between customer satisfaction, customer loyalty and customer relationship management have been created and deployed. Over time, emphasis has shifted away from raw satisfaction measures towards loyalty and CRM. The most sophisticated measurement in this area demands both technical expertise and large quantities of data.

Overall challenges

Increasingly, marketers need to consider how to move from measures to measurement systems that strategically align marketing activities with critical outcomes. This is complicated by the need to reconcile multiple purposes and multiple measures for different decision makers inside and outside the organization. The role of short- and long-term feedback from performance to activities remains complicated and poorly calibrated.

Conclusions

A thousand flowers are blooming in marketing performance measurement; thanks partly to external pressures to measure, there is no longer any difficulty in attracting the attention of either executives or scholars to measurement. Indeed, to some extent measures are no longer the problem: we have many measures available for almost any marketing phenomenon one could examine. Looking ahead, the larger issues lie in understanding how these measures are interlinked and in determining how to use them effectively in management.

References

- Aaker, D.A., and Jacobson, R. (1994). The financial information content of perceived quality. *Journal of Marketing Research*, 31(2), 191–201.
- Abela, A.V., Clark, B.H., and Ambler, T. (2005). The impact of marketing performance assessment and marketing dashboards on firm performance and learning. Working paper, Washington, DC: Catholic University of America.
- Ailawadi, K.L., Lehmann, D.R., and Neslin, S.A. (2003). Revenue premium as an outcome measure of brand equity. *Journal of Marketing*, 67(4), 1–17.
- Ambler, T. (2003). *Marketing and the Bottom Line*, 2nd edn. London: Financial Times – Prentice Hall.
- Ambler, T., and Barwise, P. (1998). The trouble with brand valuation. *The Journal of Brand Management*, 5(5), 367–77.
- Ambler, T., and Riley, D. (2000). *Marketing Metrics: A Review of Performance Measures in Use in the UK and Spain*, Report no.00–500. Cambridge, MA: Marketing Science Institute.
- American Marketing Association (2005). *Marketing Accountability Study White Paper*. Chicago: American Marketing Association.
- Anderson, E.W., Fornell, C., and Mazvancheryl, S.K. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, 68(4), 172–85.

- Anderson, E.W., and Sullivan, M.W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, 12(2), 125–43.
- Armstrong, J.S., and Collopy, F. (1996). Competitor orientation: effects of objectives and information on managerial decisions and profitability. *Journal of Marketing Research*, 33(2), 188–99.
- Bacon, C. (2002). Promo optimizer: Colgate-Palmolive approach to maximizing ROI. In Bahadir, S.C., and Tuli, K.R. (eds.), *Measuring Marketing Productivity: Linking Marketing to Financial Returns*, Report no. 02–119, 31–2. Cambridge, MA: Marketing Science Institute.
- Bahadir, S.C., and Tuli, K.R. (2002). *Measuring Marketing Productivity: Linking Marketing to Financial Returns*, Report no. 02–119. Cambridge, MA: Marketing Science Institute.
- Barry, T.E. (2002). In defense of the hierarchy of effects: a rejoinder to Weilbacher. *Journal of Advertising Research*, 42(3), 44–7.
- Barwise, P. (1993). Brand equity: snark or boojum? *International Journal of Research in Marketing*, 10, 93–104.
- Behn, R.D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.
- Bhargava, M., Dubelaar, C., and Ramaswami, S. (1994). Reconciling diverse measures of performance: a conceptual framework and test of a methodology. *Journal of Business Research*, 31, 235–46.
- Bolton, R.N., Lemon, K.N., and Verhoef, P.C. (2004). The theoretical underpinnings of customer asset management: a framework and propositions for research. *Journal of the Academy of Marketing Science*, 32(3), 271–92.
- Bonoma, T.V., and Clark, B.H. (1988). *Marketing Performance Assessment*. Boston: Harvard Business School Press.
- Bruno, H.A., Parthasarathi, U., and Singh, N. (eds.) (2005). *Does Marketing Measure Up? Performance Metrics: Practices and Impacts*, Report no. 05–301. Cambridge, MA: Marketing Science Institute.
- Buzzell, R.D., and Chussil, M.J. (1985). Managing for tomorrow. *Sloan Management Review*, 26(4), 3–14.
- Buzzell, R.D., and Gale, B.T. (1987). *The PIMS Principles: Linking Strategy to Performance*. New York: Free Press.
- Chartered Institute of Marketing (2001). *Marketing Effectiveness*. London: Chartered Institute of Marketing.
- (2005). What is marketing? CIM's definition. London: Chartered Institute of Marketing, www.cim.co.uk/dev/html/marwha.cfm (accessed 11 August 2005).
- Clark, B.H. (1999). Marketing performance measures: history and interrelationships. *Journal of Marketing Management*, 15(8), 711–32.
- (2000). Managerial perceptions of marketing performance: efficiency, adaptability, effectiveness and satisfaction. *Journal of Strategic Marketing*, 8(1), 3–25.
- (2002). Measuring performance: the marketing perspective. In Neely, A.D. (ed.), *Business Performance Measurement: Theory and Practice*. Cambridge: Cambridge University Press.
- Clark, B.H., Abela, A.V., and Ambler, T. (2004). Return on measurement: relating marketing metrics practices to firm performance. Paper presented at American Marketing Association Summer Educators' Conference, Boston, 6–9 August.

- CMO Council (2004). *Measures and Metrics: The CMO Council Marketing Performance Measurement Report*. Palo Alto, CA: CMO Council.
- Day, G.S., and Fahey, L. (1988). Valuing market strategies. *Journal of Marketing*, 52(3), 45–57.
- Debruyne, M., and Mubbard, K. (2000). *Marketing Metrics*, Report no. 00–119. Cambridge, MA: Marketing Science Institute.
- DeChernatony, L., Riley, F.D.O., and Harris, F. (1998). Criteria to assess brand success. *Journal of Marketing Management*, 14, 765–81.
- Dick, A.S., and Basu, K. (1994). Customer loyalty: toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22(2), 99–113.
- Dickson, P.R., Farris, P.W., and Verbeke, W.J.M.I. (2001). Dynamic strategic thinking. *Journal of the Academy of Marketing Science*, 29(3), 216–37.
- Donaher, P.J., and Mattson, J. (1994). Customer satisfaction during the service delivery process. *European Journal of Marketing*, 28(5), 5–16.
- Doyle, S. (2004). Which part of my marketing spends really works? Marketing mix modelling may have an answer. *Database Marketing and Customer Strategy Management*, 11(4), 379–85.
- Duffy, M. (2000). Kraft's return on marketing investment: portfolio planning implications. In Debruyne, M., and Hubbard, K. (eds.), *Marketing Metrics*, Report no. 00–119, 33–6. Cambridge, MA: Marketing Science Institute.
- (2002). Measuring marketing ROI...Great, but usually late. In Bahadir, S.C., and Tuli, K.R. (eds.), *Measuring Marketing Productivity: Linking Marketing to Financial Returns*, Report no. 02–119, 23–5. Cambridge, MA: Marketing Science Institute.
- East, R., Lomax, W., and Narain, R. (2001). Customer tenure, recommendation and switching. *Journal of Customer Satisfaction, Dissatisfaction and Complaining Behavior*, 14, 46–54.
- Echambadi, N.V. (2005). Unraveling the marketing mystique. *Strategic Finance*, 87(1), 41–7.
- Feder, R.A. (1965). How to measure marketing performance. *Harvard Business Review*, 43(3), 132–42.
- Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., and Bryant, B.E. (1996). The American customer satisfaction index: nature, purpose, and findings. *Journal of Marketing*, 60(4), 7–18.
- Franses, P.H. (2005). On the use of econometric models for policy simulation in marketing. *Journal of Marketing Research*, 52(1), 4–14.
- Geyskens, I., Gielens, K., and Dekimpe, M.G. (2002). The market valuation of internet channel additions. *Journal of Marketing*, 66(2), 102–19.
- Goodman, S.J. (1970). *Techniques of Profitability Analysis*. New York: Wiley-Interscience.
- (1972). *The Marketing Controller*. New York: AMR International.
- Greene, W.H. (1990). *Econometric Analysis*. New York: Macmillan.
- Greve, H. (1998). Performance, aspirations, and risky organizational change. *Administrative Science Quarterly*, 43(1), 58–86.
- Gruca, T.S., and Rego, L.L. (2005). Customer satisfaction, cash flow, and shareholder value. *Journal of Marketing*, 69(3), 115–30.
- Gunasekaran, A., Patel, C., and McGaughey, R.E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, 87, 333–47.

- Gupta, S., and Lehmann, D.R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17(1), 9–24.
- Gupta, S., Lehmann, D.R., and Stuart, J.A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.
- Haigh, D. (1998). *The Future of Brand Value Reporting*. London: Brand Finance.
- Haigh, D., and Knowles, J. (2004). Don't waste time with brand valuation. *Marketing NPV*, 1(6), 17–19.
- Halstead, D., Hartman, D., and Schmidt, S.L. (1994). Multisource effects on the satisfaction formation process. *Journal of the Academy of Marketing Science*, 22(2), 114–29.
- Hanssens, D. (2005). Performance regimes and marketing policy shifts. In Bruno, H.A., Parthasarathi, U., and Singh, N. (eds.), *Does Marketing Measure Up? Performance Metrics: Practices and Impacts*, Report no. 05–301, 6–7. Cambridge, MA: Marketing Science Institute.
- Henderson, B.D. (1973). *The Experience Curve—Reviewed: IV. The Growth Share Matrix or the Product Portfolio*, Perspectives 135. Boston: Boston Consulting Group.
- Hogan, J.E., Lehmann, D.R., Merino, M., Srivastava, R.K., Thomas, J.S., and Verhoef, P.C. (2002). Linking customer assets to financial performance. *Journal of Service Research*, 5(1), 26–38.
- Honeycutt, E.D. (2005). Technology improves sales performance – doesn't it? An introduction to the special issue on selling and sales technology. *Industrial Marketing Management*, 34(4), 301–4.
- Humby, C. (2005). Leveraging returns from a \$100m investment in customer rewards, through communication, price, and promotion. In Bruno, H.A., Parthasarathi, U., and Singh, N. (eds.), *Does Marketing Measure Up? Performance Metrics: Practices and Impacts*, Report no. 05–301, 4–6. Cambridge, MA: Marketing Science Institute.
- Jacobson, R. (1988). Distinguishing among competing theories of the market share effect. *Journal of Marketing*, 52(4), 68–80.
- Kaplan, R.S., and Norton, D.P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.
- (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 74(1), 75–85.
- Keefe, L.M. (2004). What is the meaning of “marketing”? *Marketing News*, 38(15), 17–18.
- Keller, K.L. (1993). Conceptualizing, measuring, and managing customer-based equity. *Journal of Marketing*, 57(1), 1–22.
- (2002). *Strategic Brand Management*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.
- Kerin, R.A., and Sethuraman, R. (1998). Exploring the brand value–shareholder value nexus for consumer goods companies. *Journal of the Academy of Marketing Science*, 26(4), 260–73.
- Kerssens-van Drongelen, I., Nixon, B., and Pearson, A. (2000). Performance measurement in industrial R&D. *International Journal of Management Reviews*, 2(2), 111–43.
- Kim, J., and Morris, J.D. (2003). The effect of advertising on the market value of firms: empirical evidence from the Super Bowl ads. *Journal of Targeting, Measurement and Analysis for Marketing*, 12(1), 53–65.
- Kim, J., Suh, E., and Hwang, H. (2003). A model for evaluating the effectiveness of CRM using the balanced scorecard. *Journal of Interactive Marketing*, 17(2), 5–19.

- Kincaid, S. (2000). Pushing the levers that make money: econometrics and business development. In Debruyne, M., and Hubbard, K. (eds.), *Marketing Metrics*, Report no. 00–119, 15–16. Cambridge, MA: Marketing Science Institute.
- Kumar, V., and Petersen, J.A. (2004). Maximizing ROI or profitability. *Marketing Research*, 16(3), 28–34.
- Lane, V., and Jacobson, R. (1995). Stock market reactions to brand extension announcements: the effects of brand attitude and familiarity. *Journal of Marketing*, 59(1), 63–77.
- Lant, T.K. (1992). Aspiration level adaptation: an empirical exploration. *Management Science*, 38(5), 623–45.
- Lee, R.P., and Grewal, R. (2004). Strategic responses to new technologies and their impact on firm performance. *Journal of Marketing*, 68(4), 157–71.
- Lehmann, D.R. (2004). Metrics for making marketing matter. *Journal of Marketing*, 68(4), 73–5.
- Lenskold, J.D. (2004). Customer-centric marketing ROI. *Marketing Management*, 13(1), 26–31.
- Lewin, A.Y., and Minton, J.W. (1986). Determining organizational effectiveness: another look, and an agenda for research. *Management Science*, 32(5), 514–38.
- Manchanda, P., Rossi, P.E., and Chintagunta, P.K. (2004). Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, 41(4), 467–78.
- March, J.G., and Sutton, R.I. (1997). Organizational performance as a dependent variable. *Organization Science*, 8(6), 698–706.
- Marketing Leadership Council (2001). *Measuring Marketing Performance*. Washington, DC: Marketing Leadership Council.
- (2002). *Marketing Investment and Impact Benchmarks*. Washington, DC: Marketing Leadership Council.
- (2003). *2003 Marketing Investment Benchmarks*. Washington, DC: Marketing Leadership Council.
- Marketing Science Institute (1998). *1998–2000 Research Priorities: A Guide to MSI Research Programs and Procedures*. Cambridge, MA: Marketing Science Institute.
- (2000). *2000–2002 Research Priorities: A Guide to MSI Research Programs and Procedures*. Cambridge, MA: Marketing Science Institute.
- (2002). *2002–2004 Research Priorities: A Guide to MSI Research Programs and Procedures*. Cambridge, MA: Marketing Science Institute.
- (2004). *2004–2006 Research Priorities: A Guide to MSI Research Programs and Procedures*. Cambridge, MA: Marketing Science Institute.
- Mavrinac, S., and Siesfeld, T. (1997). Measures that matter: an exploratory investigation of investors' information needs and value priorities. Working paper, Ivey School of Business, University of Western Ontario, London, ON, Canada.
- Menon, A., and Varadarajan, P.R. (1992). A model of marketing knowledge use within firms. *Journal of Marketing*, 56(4), 53–71.
- Miyazaki, A.D. (2001). Assessing market value of event sponsoring: corporate Olympic sponsorships. *Journal of Advertising Research*, 41(1), 9–15.
- Miller, D. (1994). What happens after success: the perils of excellence. *Journal of Management Studies*, 31(3), 325–58.

- Mittal, V., and Kamakura, W.A. (2001). Satisfaction, repurchase intent, and repurchase behavior: investigating the moderating effects of customer characteristics. *Journal of Marketing Research*, 38(1), 131–42.
- Mizik, N., and Jacobson, R. (2003). Trading off between value creation and value appropriation: the financial implications of shifts in strategic emphasis. *Journal of Marketing*, 67(1), 63–76.
- Morgan, N.A., Anderson, E.W., and Mittal, V. (2005). Understanding firms' customer satisfaction information usage. *Journal of Marketing*, 69(3), 131–51.
- Morgan, N.A., Clark, B.H., and Gooner, R. (2002). Marketing productivity, marketing audits, and systems for marketing performance assessment: integrating multiple perspectives. *Journal of Business Research*, 55, 363–75.
- Neely, A. (1998). *Measuring Business Performance*. London: Economist Books.
- Neff, J. (2005). ROI: the marketer's obsession. *Advertising Age*, 76(25), S1–S2.
- Netemeyer, R.G., Krishnan, B., Pullig, C., Wang, G., Yagci, M., Dean, D., Ricks, J., and Wirth, F. (2004). Developing and validating measures of facets of customer-based brand equity. *Journal of Business Research*, 57, 209–24.
- Piercy, N. (1986). Marketing asset accounting: scope and rationale. *European Journal of Marketing*, 20(1), 5–15.
- (2002). *Market-Led Strategic Change*, 3rd edn. Oxford: Butterworth-Heinemann.
- Reibstein, D., Norton, D., Joshi, Y., and Farris, P. (2005). Marketing dashboards: a decision support system for assessing marketing productivity. Paper presented at Marketing Science conference, Atlanta, 16–18 June.
- Reichheld, F.E. (1994). Loyalty and the renaissance of marketing. *Marketing Management*, 2(4), 10–21.
- Reinartz, W., Krafft, M., and Hoyer, W.D. (2004). The customer relationship management process: its measurement and impact on performance. *Journal of Marketing Research*, 41(3), 293–305.
- Reinartz, W., and Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. *Journal of Marketing*, 64(4), 17–35.
- Rigby, D. (2001). Management tools and techniques: a survey. *California Management Review*, 43(2), 139–60.
- Rust, R.T., Ambler, T., Carpenter, G.S., Kumar, V., and Srivastava, R.K. (2004). Measuring marketing productivity: current research and future directions. *Journal of Marketing*, 68(4), 76–89.
- Rust, R.T., Lemon, K.N., and Zeithaml, V.A. (2004). Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1), 109–27.
- Selnes, F. (1993). An examination of the effect of product performance on brand reputation, satisfaction and loyalty. *European Journal of Marketing*, 27(9), 19–35.
- Sevin, C.H. (1965). *Marketing Productivity Analysis*. New York: McGraw-Hill.
- Shank, J.K. (1989). Strategic cost management: new wine, or just new bottles? *Journal of Management Accounting Research*, 1, 47–65.
- Sharma, A., and Lacey, N. (2004). Linking product development outcomes to market valuation of the firm: the case of the U.S. pharmaceutical industry. *Journal of Product Innovation Management*, 21(5), 297–308.

- Sheth, J.N., and Sisodia, R.S. (1995). Feeling the heat. *Marketing Management*, 4(2), 8–23.
- Simon, C.J., and Sullivan, M.W. (1993). The measurement and determinants of brand equity: a financial approach. *Marketing Science*, 12(1), 28–52.
- Spreng, R.A., MacKenzie, S.B., and Olshavsky, R.W. (1996). A reexamination of the determinants of customer satisfaction. *Journal of Marketing*, 60(3), 15–32.
- Srivastava, R.K., Shervani, T.A., and Fahey, L. (1998). Market-based assets and shareholder value: a framework for analysis. *Journal of Marketing*, 62(1), 2–18.
- (1999). Marketing, business processes, and shareholder value: an organizationally embedded view of marketing activities and the discipline for marketing. *Journal of Marketing*, 63 (Special issue), 168–79.
- Szymanski, D.M., Bharadwaj, S.G., and Varadarajan, P.R. (1993). An analysis of the market share–profitability relationship. *Journal of Marketing*, 57(3), 1–18.
- Szymanski, D.M., and Henard, D.H. (2001). Customer satisfaction: a meta-analysis of the evidence. *Journal of the Academy of Marketing Science*, 29(1), 16–35.
- Teas, R.K. (1993). Expectations, performance evaluation, and consumers’ perceptions of quality. *Journal of Marketing*, 57(4), 18–34.
- Teas, R.K., and Palan, K.M. (1997). The realms of scientific meaning framework for constructing theoretically meaningful nominal definitions of marketing concepts. *Journal of Marketing*, 61(2), 52–67.
- Toulson, P.K., and Dewe, P. (2004). HR accounting as a measurement tool. *Human Resource Management Journal*, 14(2), 75–90.
- Vazquez, R., Del Rio, A.B., and Iglesias, V. (2002). Consumer-based brand equity: development and validation of a measurement scale. *Journal of Marketing Management*, 18(1/ 2), 27–48.
- Venkatesan, R., and Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4), 106–25.
- Verhoef, P.C., Spring, P.N., Hoekstra, J.C., and Leeftang, P.S.H. (2003). The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. *Decision Support Systems*, 34(4), 471–82.
- Voss, G.B., Parasuraman, A., and Grewal, D. (1998). The roles of price, performance, and expectations in determining satisfaction in service exchanges. *Journal of Marketing*, 62(4), 46–61.
- Webster, F.E. Jr., Malter, A.J., and Ganesan, S. (2005). The decline and dispersion of marketing competence. *Sloan Management Review*, 46(4), 35–43.
- Weilbacher, W.M. (2001). Point of view: does advertising cause a “hierarchy of effects”? *Journal of Advertising Research*, 41(6), 19–26.
- Winer, R.S. (2000). What marketing metrics are used by MSI members? In Debruyne, M., and Hubbard, K. (eds.), *Marketing Metrics*, Report no. 00–119, 3–6. Cambridge, MA: Marketing Science Institute.
- Yoo, B., and Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, 52, 1–14.
- Yi, Y. (1990). A critical review of customer satisfaction. In Zeithaml, V.A. (ed.), *Review of Marketing 1990*, 68–123. Chicago: American Marketing Association.

3

Measuring performance: the operations management perspective

Andy Neely

Introduction

This chapter explores performance measurement from an operations perspective. Members of the operations management community have been interested in performance measurement – at both the strategic and the tactical level – for decades. Wickham Skinner, widely credited with the development of the operations strategy framework, for example, observed in 1974:

A factory cannot perform well on every yardstick. There are a number of common standards for measuring manufacturing performance. Among these are short delivery cycles, superior product quality and reliability, dependable delivery promises, ability to produce new products quickly, flexibility in adjusting to volume changes, low investment and hence higher return on investment, and low costs.

These measures of manufacturing performance necessitate trade-offs – certain tasks must be compromised to meet others. They cannot all be accomplished equally well because of the inevitable limitations of equipment and process technology. Such trade-offs as costs versus quality or short delivery cycles versus low inventory investment are fairly obvious. Other trade-offs, while less obvious, are equally real. They involve implicit choices in establishing manufacturing policies. (Skinner, 1974, 115)

Skinner's early work on operations (initially manufacturing) strategy influenced a generation of researchers, all of whom subsequently explored the question of how to align operations policies with operations objectives (see, for example, Hayes and Wheelwright, 1984; Hayes, Wheelwright and Clark, 1988; Hill, 1985; Mills, Platts and Gregory, 1995; Platts and Gregory, 1990; Slack and Lewis, 2001).

It was during this phase of exploration that the operations management community's interest in performance measurement was probably at its peak, with numerous authors asking how organizational performance measurement systems could be aligned with operations strategies (for a comprehensive review of these works, see Neely, Gregory and Platts, 1995). Initially interest was sparked by the growing recognition – in both the accounting and

operations management communities – that the performance measurement systems used in many firms were hindering efforts to improve manufacturing performance. In 1980, for example, Hayes and Abernathy wrote a provocatively entitled paper – “Managing our way to economic decline” – in which they argued that the performance measurement and incentive schemes used in many US firms resulted in a short-term orientation that was significantly damaging the United States’ ability to compete internationally (Hayes and Abernathy, 1980). A few years later Turney and Anderson published an influential paper in the *Sloan Management Review*, which highlighted the problems that traditional cost accounting systems posed for firms that were seeking to change the way they worked by adopting new manufacturing methods, most notably those based on the just-in-time concept (Turney and Anderson, 1989).

Clearly, as time has passed the operations management community’s interest in performance measurement has matured. Today there is a greater emphasis on exploring the performance of service operations, especially in a global and networked context. The community’s understanding of performance trade-offs is more sophisticated, and it is now widely recognized that trade-offs are time- and context-dependent. Hence a key question for scholars of operations management is establishing how organizations can build cumulative capabilities that mitigate the impact of performance trade-offs (interested readers should see work on path dependency models for a fuller summary, such as Ferdows and De Meyer, 1990, Miller and Roth, 1994, and Flynn and Flynn, 2004).

The remainder of this chapter explores these themes and developments in more detail, by reviewing research on performance measurement conducted by members of the operations management community. The chapter consists of three broad sections. The first reviews operations management prior to its resurgence in the early 1980s. The second section explores the evolution of performance measurement research between the 1980s and the current time. The third poses some challenges for the future of operations management research on performance measurement.

Operations as a cost centre

In the period immediately following the Second World War demand for manufactured goods outstripped supply in the vast majority of industries, with many countries effectively having to rebuild their manufacturing

capacity. Given these capacity constraints, organizations found that they could sell virtually everything they made (Deming, 1982, 1986). Hence their primary focus was on making a narrow range of products as cheaply as possible and then selling them to relatively undemanding customers; remember Henry Ford's famous line about the Model T: "You can have it in any colour as long as it is black." Given the broader economic context, the primary role of the operations function was to produce goods as rapidly and as efficiently as possible. Hence, the question that operations managers wanted their measurement systems to help them answer was: "How efficient are we?"

The result was a stream of research on productivity measurement and management (Burgess, 1990; Kendrick, 1984; Sink, 1985). Some authors concentrated on contrasting different dimensions of productivity. Bicheno (1989), for example, explored the different ways in which productivity can be enhanced. Ruch (1982), on the other hand, asked how traditional measures of blue-collar productivity could be used in white-collar environments. A particularly significant stream of literature was that associated with the measurement of total factor productivity. The thesis underpinning this literature was that too often managers relied on partial measures of productivity – e.g. labour productivity. Several authors questioned the veracity of single-dimensional measures of productivity on the grounds that they provided only a partial picture of firm performance (Craig and Harris, 1973; Mundel, 1987). In reality, the productivity a firm achieves is a function of how efficiently it uses all its inputs – labour, capital, technology and energy – to produce outputs (Hayes and Clark, 1986).

Clearly, this work on productivity mirrors much of the work undertaken by other communities, most notably by economists (O'Mahony and de Boer, 2002). However, from an operations perspective, the focus was very much within the walls of the factory.

A strategic perspective on operations

By the late 1970s operations management, as an academic discipline, was in some turmoil. Influential business schools in the United States had begun to question whether they needed operations management departments, and the field was struggling to define its identity (Buffa, 1980; Chase, 1980). This crisis in confidence was the result of people leaving

the field of operations management to establish new functional fields with their own identities. Indeed, Buffa argued that, in the mid-1950s, production/operations management was effectively synonymous with the entire field of industrial management, and that elements from functional fields such as finance, marketing and human resources were all encompassed under the production/operations management umbrella. As these functional fields developed the production/operations management community began to disintegrate, and those who remained found themselves fighting for the survival of a discipline that had been stripped of all but a few techniques, such as “time and motion study, plant layout, Gantt’s production control boards, the simple economic order quantity (EOQ) model and simplistic descriptions of how production systems worked” (Buffa, 1980, 1).

Throughout the 1960s and 1970s operations research was effectively the saviour of operations management. Operations research techniques, such as linear programming and data envelopment analysis (DEA), were applied to pragmatic operations management problems such as lot sizing, scheduling and production control (Charnes, Cooper and Rhodes, 1978; Buffa, 1980; Chase, 1980; Banker, Charnes and Cooper, 1984). Indeed, a review of the first twenty-five years of management science found that operations management problems were the most commonly studied (Chase, 1980). By the mid-1970s the field of production/operations management was again in crisis, however, as the operations research community began to apply their tools and techniques to other fields of management. What saved production/operations management this time was the emergence of Japan as an economic power and the widespread soul-searching that resulted in the United States and Europe as managers and academics asked themselves what underpinned the Japanese economic miracle. Numerous answers – many of them partial – were offered. But then authors began to propose some more rounded insights. Schonberger (1982, 1986) and Hall (1983), for example, argued that the operations in Japanese firms were simply better managed. Monden (1996) documented the rationale and workings of the Toyota production system, while Imai (1986) highlighted the importance of kaizen. The 1980s also saw a rise in the popularity of the “quality gurus”, most notably Crosby, Deming, Feigenbaum and Juran. Together, these developments resulted in a resurgence of interest in the measurement of operations performance, especially in terms of the five operations performance objectives – quality, dependability, speed, cost and flexibility – originally proposed by Skinner (1969).

The five operations performance objectives

These five operations performance objectives underpin much of the work on performance measurement that has been undertaken subsequently by members of the operations management community. There are several points to note about these performance objectives. The first is that they are all multidimensional. “Quality” is not simply a reference to conformance to specification, but also encompasses a variety of other dimensions – e.g. performance (how well the product performs its primary function), features (what additional features the product contains), reliability (how well the product continues to perform), technical durability (how long the product lasts before becoming technically obsolete), serviceability (how easy the product is to service), aesthetics (the look and feel of the product), perceived quality (the customer’s perception of the product’s quality) and value for money (Garvin, 1987; Schonberger, 1990; Neely and Wilson, 1992). Similarly, “speed” can refer to the time taken to generate quotes, the time taken to deliver (delivery speed), the frequency with which deliveries can be made (delivery frequency), the time taken to produce the product (production speed) and the time taken to develop new products (Stalk, 1988), while “dependability” can refer to schedule adherence (keeping to plan), delivery performance (is the product delivered on time in full?) and price performance – indeed, the general ability to meet promises.

The most multidimensional of the five performance objectives is that of flexibility. Slack makes a useful distinction between range flexibility and response flexibility (Slack, 1983, 1987). Range flexibility is the ability to cope with a wide range of requirements, while response flexibility is the ability to change quickly. As well as these two dimensions of flexibility, authors have also identified different forms of flexibility (Gerwin, 1987; Slack, 1987) – e.g. material quality (the ability to cope with incoming materials of varying quality), output quality (the ability to produce outputs of varying quality), new product flexibility (the ability to cope with the introduction of new products), product modification (the ability to cope with modified products), deliverability (the ability to cope with changed delivery schedules), volume (the ability to cope with changed production volumes), mix (the ability to cope with different production mixes) and resource mix (the ability to cope with different resource mixes). Figure 3.1 summarizes these different dimensions, while the subsections that follow explore them in more detail.

Quality	Dependability	Speed	Flexibility
Q1: Performance	D1: Schedule adherence	S1: Quote generation	F1: Material quality
Q2: Features	D2: Delivery	S2: Delivery speed	F2: Output quality
Q3: Reliability	D3: performance	S3: Delivery frequency	F3: New product
Q4: Conformance	D3: Price performance	S4: Production speed	F4: Modified product
Q5: Technical durability	D4: Ability to keep promises	S5: New product development speed	F5: Deliverability
Q6: Serviceability			F6: Volume
Q7: Aesthetics			F7: Mix
Q8: Perceived quality			F8: Resource mix
Q9: Value for money			
		Cost	
		C1: Manufacturing cost	
		C2: Value added	
		C3: Selling price	
		C4: Running cost	
		C5: Service cost	
		C6: Profit	

Figure 3.1: The multiple dimensions of the five operations performance objectives

Before moving on to explore the five operations performance objectives in more detail it is worth making two additional points. The first is that the five operations performance objectives trade off with one another, but that the extent of these trade-offs are time- and context-specific. High quality, defined in terms of product performance, can be delivered, but potentially at a cost. Tight delivery schedules can be met, but perhaps only by investing in additional resource. Operations managers constantly strive to find ways of pushing back the performance frontiers of these five performance objectives by enhancing their operation's capabilities so that the impact of the trade-offs can be mitigated over time. The continuous search for enhanced operational capabilities underpins, for example, the manufacturing philosophy of kaizen – continuous improvement of the operation in a search for new and better ways of delivering performance.

The second point to note about the five operations performance objectives is that there are internal as well as external reasons why organizations might wish to excel at them (Slack, 1991). Higher quality, for example, allows the organization to deliver higher-quality (and potentially more valuable) products. Higher quality, however, also means fewer mistakes, hence less rework, which is costly to the operation if it has to be performed. Speed means that the organization can respond to customer requests more rapidly, but also means that the organization's capital is tied up for shorter periods of time in the form of inventories and work in progress. The same argument can be made for each of the five operations performance objectives. Hence, one of the key strategic challenges for operations managers is to decide on which of the sub-dimensions of these five performance objectives they wish their operation to excel and how they are going to configure the operation to do so.

The measurement of operations performance: quality

As already discussed, quality was traditionally defined in terms of conformance to specification. Hence quality-based measures of performance focused on issues such as the number of defects produced and the cost of quality. Feigenbaum (1961) was the first to suggest that the true cost of quality is a function of three types of quality cost: prevention, appraisal and failure costs. Campanella and Corcoran (1983) offer the following as definitions of these three types of cost.

- (1) Prevention costs are those costs expended in an effort to prevent discrepancies, such as the costs of quality planning, supplier quality surveys and training programmes.
- (2) Appraisal costs are those costs expended in the evaluation of product quality and in the detection of discrepancies, such as the costs of inspection, test and calibration control.
- (3) Failure costs are those costs expended as a result of discrepancies, and are usually divided into two types: (i) internal failure costs are costs resulting from discrepancies found prior to delivery of the product to the customer, such as the costs of rework, scrap and material review; and (ii) external failure costs are costs resulting from discrepancies found after delivery of the product to the customer, such as the costs associated with the processing of customer complaints, customer returns, field services and warranties.

Crosby's assertion (1972) that "quality is free" is based on the assumption that, for most firms, an increase in prevention costs will be more than offset by a decrease in failure costs. Plunkett and Dale (1988) point out that, although conceptually appealing, the academic rigour underpinning the cost of quality model is debatable. It is based on assumptions and estimates, rather than on data. And, like the EOQ model, it is questionable whether an optimum level of quality really exists.

With the advent of total quality management (TQM), the emphasis shifted away from "conformance to specification" and towards customer satisfaction. As a result, the use of customer opinion surveys and market research became much more widespread. The establishment of the Malcolm Baldrige National Quality Award in the United States and the European Quality Award reflected this trend. Other common measures of quality include statistical process control (Crosby, 1972; Deming, 1982, 1986) and the Motorola six-sigma concept (which has gained significant support in recent years – see, for example, Pande, Neuman and Cavanagh, 2001). These last

two measures of quality raise an important issue relevant to performance measurement for operations, because they focus on the measurement of the process rather than the output (a theme that is gaining increasing support in the performance measurement literature).

The measurement of operations performance: speed and dependability

Time has been described as both a source of competitive advantage and the fundamental measure of operations performance (Stalk, 1988). The “just-in-time” (JIT) manufacturing philosophy is based on the assumption that the production or delivery of goods either too early or too late is wasteful (Monden, 1996). Other manufacturing planning techniques, such as optimized production technology (OPT), seek the minimization of throughput times (Goldratt and Cox, 2004). Indeed, OPT was the stimulus for Galloway and Waldron (1988a, 1988b, 1989a, 1989b) to develop the notion of throughput accounting, which is based on the assumption that throughput in operations is constrained by bottleneck processes. Hence, the key to understanding and managing profitability is to identify resource constraints (bottleneck processes) and ensure that these are fully utilized. Of course, in reality, it proves difficult to identify resource constraints, not least because these can be affected by the sales volume and variety at any point in time.

The measurement of operations performance: cost

In addition to the operations management community’s work on productivity (discussed previously), there is also widespread interest in product costing. The developments that the operations management community refer to in this area are frequently the result of work in the accounting community. Johnson and Kaplan’s *Relevance Lost* (Johnson and Kaplan, 1987) and Johnson’s broader works on the evolution of management accounting systems, for example, are widely cited (Johnson, 1972, 1975a, 1975b, 1978, 1980, 1981, 1983), as is Kaplan and Cooper’s work on activity-based costing (Kaplan and Cooper, 1997). Mainstream operations management academics have also, however, contributed to these debates. Miller and Vollmann (1985), for example, talk about the hidden costs of factories. They argue that many managers focus on the visible costs – e.g. direct labour and material – while the majority of overheads are caused by the “invisible” transaction costs. These observations clearly marry with the

assumptions underpinning activity-based costing, especially the notion that it is activities, and not products, that cause cost.

The measurement of operations performance: flexibility

One of the most substantive areas of debate in the operations management community is in terms of the fifth operations performance objective: flexibility. Slack (1983) identifies range, cost and time as dimensions of flexibility, although he later modifies this model so that it includes only range and response, where range refers to the issue of how far the manufacturing system can change and response focuses on the question of how rapidly and cheaply it can change (Slack, 1987).

Gerwin (1987) observes that very little is known about the implications of flexibility for manufacturing management, and suggests that “part of the problem arises from the lack of operational measures of flexibility”. After identifying various dimensions of flexibility he suggests the following measures.

- Mix flexibility – measured in terms of the number of components handled by the equipment.
- Changeover flexibility – measured in terms of the number of component substitutions made over a given time period.
- Modification flexibility – measured in terms of the number of design changes made to a component per time period.
- Rerouting flexibility – measured in terms of lost production when a machine breaks down.
- Volume flexibility – measured in terms of the average volume fluctuations that can be coped with over a given time period divided by the capacity limit.
- Material flexibility – measured in terms of the extent to which variations in key dimensional and metallurgical properties can be handled by equipment.
- Sequencing flexibility – measured in terms of the number of different sequences handled by the equipment.

The need for methodologies and frameworks

Throughout the 1980s and 1990s the work on measuring the operations performance objectives was accompanied by a growing interest in Skinner’s

early work on manufacturing strategy (Skinner, 1969). In both the United States and the United Kingdom conferences devoted to manufacturing strategy were held in the early 1980s. The resultant debates highlighted how important it was for the operations management community to consider how to align performance measurement systems with operations strategies (Neely *et al.*, 1994). Interestingly, the themes raised in these debates echoed some earlier writings – illustrating the enduring nature of some performance measurement challenges. Ridgway, writing in the second edition of the *Administrative Science Quarterly* half a century ago, had published a paper entitled “Dysfunctional consequences of performance measurements” (Ridgway, 1956). In that paper Ridgway explored the relative strengths and weaknesses of single, multiple and aggregated performance measures, bemoaning the “strong tendency to state numerically as many as possible of the variables with which management must deal”. A few years earlier – in 1952 – Chris Argyris, in his classic text *The Impact of Budgets on People*, had reported that managers claimed to “feed machines all the easy orders at the end of the month to meet [their] quota” (Argyris, 1952). These two themes – *the desire to quantify* and *the unanticipated consequences of measurement* – led that doyen of management, Peter Drucker, to argue that one potential solution was to introduce “balanced” sets of measures. “Market standing, innovation, productivity, physical and financial resources, profitability, manager performance and development, worker performance and attitude, and public responsibility” are appropriate performance criteria, said Drucker in his 1954 publication *The Practice of Management* (Drucker, 1954).

If the clock is turned forward fifty years, we find the same themes still being discussed (and not only in the operations management literature). Power’s book *The Audit Society: Rituals of Verification* bemoans the rise of the “audit society”, arguing that practitioners and policy makers have become obsessed with measurement and regulation (Power, 1997) – the desire to quantify. As already mentioned, Hayes and Abernathy explore the unintended consequences of this obsession in their paper “Managing our way to economic decline” (Hayes and Abernathy, 1980). Johnson and Kaplan expand these arguments, claiming that not only did measurement systems result in unintended consequences, but also that the measurement systems many firms used were woefully inadequate because they provided managers with redundant information as they were based on assumptions that were grossly outdated, given the changing nature of organizational cost structures (Johnson and Kaplan, 1987). This is a point reinforced by Chandler in his book *The Visible Hand*, which emphasizes that many of the

basic principles of accounting remained largely unchanged from when they were first developed in the 1920s by the DuPont cousins and Donaldson Brown (Chandler, 1977).

These recurring themes – *the desire to quantify* and *the unanticipated consequences of quantification* – appear to have resulted in frequent “rediscoveries” of Drucker’s 1954 suggestion that balanced measurement systems should be developed (Drucker, 1954). Throughout the 1980s and early 1990s numerous authors suggested measurement frameworks that might be appropriate – the performance pyramid (Lynch and Cross, 1991), the results-determinants framework (Fitzgerald *et al.*, 1991), the performance measurement matrix (Keegan, Eiler and Jones, 1989), the balanced scorecard (Kaplan and Norton, 1992) and the performance prism (Neely, Adams and Kennerley, 2002). The result was that a dominant research question in the mid-1990s, at least for the operations management community with an interest in performance measurement, was how these so-called “balanced performance measurement systems” could be developed and deployed. There followed a rich stream of work on the design and deployment of performance measurement systems, which reported on research to develop processes for designing measurement systems and barriers to their successful implementation (Bourne *et al.*, 2000; Neely *et al.*, 1996). At a more detailed level, authors also presented a variety of frameworks for designing individual measures of performance (Neely *et al.*, 1997; Globerson, 1985) and for assessing the effectiveness of organizational performance measurement systems (Dixon, Nanni and Vollmann, 1990).

Moving beyond the organization’s boundaries

While much of the work outlined in the previous section concentrated on issues within the boundaries of the firm, other streams of work on performance measurement moved outside the firm. For the operations management community, two of these were particularly important: benchmarking and supply chains.

Benchmarking became of widespread interest to the community in the mid- to late 1980s. There were two reasons for this. First, there was widespread interest within firms to understand their performance relative to others. Robert Camp, of Xerox, was one of the most vocal proponents of benchmarking, highlighting how it had not only enabled Xerox to understand how much they had to improve but also provided ideas about

how they might improve (Camp, 1989). The academic community saw the value of benchmarking from a different perspective: it gave them access to firm- and/or plant-level performance data that could be used in research. The result was a series of large-scale international comparative studies – most notably the Manufacturing Future Survey, the International Manufacturing Strategy Survey, the Global Manufacturing Research Group and the United Kingdom’s Best Factories Award. In-depth – within-firm or within-industry – studies also became popular. Indeed, one of the most influential academic books ever produced grew out of one such study: *The Machine that Changed the World* (Womack, Jones and Roos, 1990) was produced following a five-year study of the practices and performance of automotive plants around the world. Interestingly, the data gathered during this and similar studies enabled the academic community to return to their debates about performance trade-offs (Delbridge, Lowe and Oliver, 1995; New and Szejcowski, 1995) and the relationship between performance and practice (Hanson and Voss, 1995).

The second stream of work exploring measurement beyond the boundaries of the firm focused on the measurement of supply chains. Initially this work focused on how individual firms could assess the performance of their suppliers, reflecting a growing practitioner community interest in the supply chain management (Beamon, 1999). As chapter 4 illustrates, however, this early and somewhat narrow focus has blossomed and broadened in the last few years. Much more focus today is placed on the measurement of end-to-end supply chains, rather than the relationship between original equipment manufacturers and their first-tier suppliers (see chapter 4 for a fuller exposition).

From performance measurement to performance management

To date, much of the work on performance measurement that has been carried out by members of the operations management community has focused on the issue of performance measurement rather than performance management (Neely, 2005). Researchers have explored such questions as how performance measurement systems that align with strategies can be developed and deployed successfully (Bourne *et al.*, 2000; Neely *et al.*, 1996). However, measurement itself delivers little value until the data generated are analysed and acted upon. Work on this topic will require the operations management community to engage with a stream of relevant literatures. Clearly, the human resources research on high-performance workplaces is

relevant (Huselid, Becker and Ulrich, 2001), as is the same community's work on performance management at the individual level (Huselid, Becker and Beatty, 2005). Drawing the net more widely, however, means that the work of psychologists, such as Edwin Locke and Gary Latham on goal-setting theory (Locke and Latham, 1989), and economists, such as Michael Jensen on incentives and agency theory, has to be considered (Holmstrom and Milgrom, 1991; Jensen and Meckling, 1994). So does the work of those scholars who criticize concepts such as agency theory and transaction cost economics on the basis of the pessimistic assumptions they make about human motivation (Ghoshal and Moran, 1996). The underlying point is that there is a rich and wide canvas on which work can draw.

Whither performance management: the longer term

In thinking about future research agendas, it is important to position work in the broader societal developments that influence events. Clearly, the steady march of technology has significant implications for performance measurement research. Not only are there now multiple software packages for measuring and monitoring performance (see Marr and Neely, 2003, for a review) but there are also multiple ways of capturing data. EPOS (electronic point of sale) systems in retail outlets furnish massive amounts of data on customer shopping habits, as well as providing advanced means of managing inventories. Linking these data to enterprise resource planning (ERP) systems means that a wealth of data can be made available, although the challenge for managers is how to process these data and ensure that they are converted into a form that can help them better manage their organizations.

In spite of all of the research that has been undertaken, we actually know relatively little about what information managers use and how they use it. We know even less when one considers that the nature of many firms has changed significantly in recent years. Most large organizations are now clearly embedded in global networks. The rise of outsourcing and offshoring results in a new level of complexity for performance measurement, because now the data that one needs to manage performance have to be drawn from multiple, highly interdependent firms spread across the globe.

Additionally, especially in developed economies, it is widely recognized that firms cannot compete on the basis of cost (there is always somewhere cheaper to manufacture a product). Hence, these firms are being encouraged to compete on the basis of value added through constant innovation (Porter and Ketels, 2003; DTI, 2003). Of course, this raises an interesting question in

terms of measurement. Many would argue that innovation is an inherently uncontrolled and uncontrollable process: innovations occur due to happenstance and circumstance. So, what role does measurement (which is often seen as a control system) play in such circumstances? It is also worth considering the growing importance of services. Traditionally, researchers in the operations management community have focused on the manufacturing sector primarily. This overemphasis on a single sector is changing, but we still need to understand much more about performance measurement in non-manufacturing settings, such as services and the public and voluntary sectors.

Overall, it is clear that measurement is not going to go away. The desire to quantify that Ridgway identified fifty years ago still exists. External agencies (most notably regulators and legislators, in light of corporate scandals) are demanding that firms release more and more information about their performance. Hence, the pressure – both internal and external – for measurement to grow remains. The challenge is how to employ measurement in a positive way, ensuring that firms can grow and continue to appropriate value without succumbing to excessive control, coordination and micro-management.

References

- Argyris, C. (1952). *The Impact of Budgets on People*. New York: Controllershship Foundation.
- Banker, R.D., Charnes, A., and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–92.
- Beamon, B.M. (1999). Measuring supply chain performance. *International Journal of Operations and Production Management*, 19(3–4), 275–92.
- Bicheno, J. (1989). Exploring productivity accounting for management and strategy in manufacturing and services. *International Journal of Operations and Production Management*, 9(5), 56–68.
- Bourne, M.C.S., Mills, J.F., Wilcox, M., Neely, A.D., and Platts, K.W. (2000). Designing, implementing and updating performance measurement systems. *International Journal of Operations and Production Management*, 20(7), 754–71.
- Buffa, E.S. (1980). Research in operations management. *Journal of Operations Management*, 1(1), 1–8.
- Burgess, T.F. (1990). A review of productivity. *Work Study*, 39(1), 6–9.
- Camp, R.C. (1989). *Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance*. Milwaukee: ASQS Quality Press.
- Campanella, J., and Corcoran, F.J. (1983). Principles of quality costs. *Quality Progress*, 16(4), 16–22.
- Chandler, A.D. (1977). *The Visible Hand: Managerial Revolution in American Business*. Boston: Harvard University Press.

- Charnes, A., Cooper, W.W., and Rhodes, E. (1978). Measuring efficiency of decision-making units. *European Journal of Operations Research*, 2(6), 429–44.
- Chase, R.B. (1980). A classification and evaluation of research in operations management. *Journal of Operations Management*, 1(1), 9–14.
- Craig, C.E., and Harris, C.R. (1973). Total productivity measurement at the firm level. *Sloan Management Review*, 14(3), 13–29.
- Crosby, P.B. (1972). *Quality Is Free*. New York: McGraw-Hill.
- Delbridge, R., Lowe, J., and Oliver, N. (1995). The process of benchmarking: a study from the automotive industry. *International Journal of Operations and Production Management*, 15(4), 50–62.
- Deming, W.E. (1982). *Quality, Productivity and Competitive Position*. Cambridge, MA: MIT Press.
- (1986). *Out of the Crisis*. Cambridge, MA: MIT Press.
- Dixon, J.R., Nanni, A.J., and Vollmann, T.E. (1990). The new performance challenge: measuring operations for world-class competition. Homewood, IL: Dow-Jones Irwin.
- Drucker, P. (1954). *The Practice of Management*. New York: Harper.
- DTI (2003). *Innovation Report: Competing in the Global Economy*. London: Department of Trade and Industry.
- Feigenbaum, A.V. (1961). *Total Quality Control*. New York: McGraw-Hill.
- Ferdows, K., and De Meyer, A. (1990). Lasting improvements in manufacturing performance: in search of a new theory. *Journal of Operations Management*, 9(2), 168–85.
- Fitzgerald, L., Johnston, R., Brignall, S., Silvestro, R., and Voss, C. (1991). *Performance Measurement in Service Business*. London: Chartered Institute of Management Accountants.
- Flynn, B.B., and Flynn, E.J. (2004). An exploratory study of the nature of cumulative capabilities. *Journal of Operations Management*, 22(5), 439–57.
- Galloway, D., and Waldron, D. (1988a). Throughput accounting part 1: the need for a new language for manufacturing. *Management Accounting*, 66(10), 34–5.
- (1988b). Throughput accounting part 2: ranking products profitability. *Management Accounting*, 66(11), 34–5.
- (1989a). Throughput accounting part 3: a better way to control labour costs. *Management Accounting*, 67(1), 32–3.
- (1989b). Throughput accounting part 4: moving on to complex products. *Management Accounting*, 67(2), 40–1.
- Garvin, D.A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6), 101–9.
- Gerwin, D. (1987). An agenda for research on the flexibility of manufacturing processes. *International Journal of Operations and Production Management*, 7(1), 38–49.
- Ghoshal, S., and Moran, P. (1996). Bad for practice: a critique of the transaction cost theory. *Academy of Management Review*, 21(1), 13–47.
- Globerson, S. (1985). Issues in developing a performance criteria system for an organization. *International Journal of Production Research*, 23(4), 639–46.
- Goldratt, E.M., and Cox, J. (2004). *The Goal: A Process of Ongoing Improvement*. London: Gower.
- Hall, R.W. (1983). *Zero Inventories*. Homewood, IL: Dow-Jones Irwin.

- Hanson, P., and Voss, C. (1995). Benchmarking best practice in European manufacturing sites. *Business Process Engineering and Management Journal*, 1(1), 60–74.
- Hayes, R.H., and Abernathy, W.J. (1980). Managing our way to economic decline. *Harvard Business Review*, 58(4), 67–77.
- Hayes, R.H., and Clark, K.B. (1986). Why some factories are more productive than others. *Harvard Business Review*, 64(5), 66–73.
- Hayes, R.H., and Wheelwright, S.C. (1984). *Restoring Our Competitive Edge: Competing through Manufacturing*. New York: John Wiley.
- Hayes, R.H., Wheelwright, S.C., and Clark, K.B. (1988). *Dynamic Manufacturing: Creating the Learning Organisation*. New York: Free Press.
- Hill, T.J. (1985). *Manufacturing Strategy: The Strategic Management of the Manufacturing Function*. London: Macmillan.
- Holmstrom, B., and Milgrom, P. (1991). Multitask principal agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organisation*, 7, 24–52.
- Huselid, M., Becker, B., and Beatty, R. (2005). *The Workforce Scorecard*. Boston: Harvard Business School Press.
- Huselid, M., Becker, B., and Ulrich, D. (2001). *The HR Scorecard: Linking People, Strategy and Performance*. Boston: Harvard Business School Press.
- Imai, M. (1986). *Kaizen: The Key to Japan's Competitive Success*. New York: McGraw-Hill.
- Jensen, M.C., and Meckling, W.H. (1994). The nature of man. *Journal of Applied Corporate Finance*, 7(2), 4–19.
- Johnson, H.T. (1972). Early cost accounting for internal management control: Lyman Mills in the 1850s. *Business History Review*, 46(4), 466–74.
- (1975a). Management accounting in early integrated industry: E. I. duPont de Nemours powder company 1903–1912. *Business History Review*, 49(2), 184–204.
- (1975b). The role of history in the study of modern business enterprise. *Accounting Review*, 50(3), 444–50.
- (1978). Management accounting in an early multidivisional organisation: General Motors in the 1920s. *Business History Review*, 52(4), 490–517.
- (1980). Markets, hierarchies, and the history of management accounting. Paper presented at Third International Congress of Accounting Historians, London, 16–18 August.
- (1981). Toward a new understanding of nineteenth-century cost accounting. *Accounting Review*, 56(3), 510–18.
- (1983). The search for gain in markets and firms: a review of the historical emergence of management accounting systems. *Accounting, Organisations and Society*, 2/3, 139–46.
- Johnson, H.T., and Kaplan, R.S. (1987). *Relevance Lost: The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.
- Kaplan, R.S., and Cooper, R. (1997). *Cost and Effect: Using Integrated Cost Systems to Drive Profitability and Performance*. Boston: Harvard Business School Press.
- Kaplan, R.S., and Norton, D.P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.
- Keegan, D.P., Eiler, R.G., and Jones, C.R. (1989). Are your performance measures obsolete? *Management Accounting*, 70(12), 45–50.
- Kendrick, J.W. (1984). *Improving Company Productivity: Handbook with Case Studies*. Baltimore: Johns Hopkins University Press.

- Locke, E., and Latham, G. (1989). *A Theory of Goal Setting and Task Performance*. New York: Prentice Hall.
- Lynch, R.L., and Cross, K.F. (1991). *Measure Up! Yardsticks for Continuous Improvement*. Cambridge, MA: Blackwell Publishers.
- Marr, B., and Neely, A.D. (2003). *Balanced Scorecard Software Report*. Stamford, CT: Gartner.
- Miller, J.G., and Roth, A.V. (1994). A taxonomy of manufacturing strategies. *Management Science*, 40(3), 285–305.
- Miller, J.G., and Vollmann, T.E. (1985). The hidden factory. *Harvard Business Review*, 63(5), 142–50.
- Mills, J.F., Platts, K.W., and Gregory, M.J. (1995). A framework for the design of manufacturing strategy processes: a contingency approach. *International Journal of Operations and Production Management*, 15(4), 17–49.
- Monden, Y. (1996). *The Toyota Management System: Linking the Seven Key Functional Areas*. Cambridge, MA: Productivity Press.
- Mundel, M. (1987). *Measuring Total Productivity of Manufacturing Organisations*. White Plains, NY: Unipub-Kraus International.
- Neely, A.D. (2005). The evolution of performance measurement research: developments in the last decade and a research agenda for the next. *International Journal of Operations and Production Management*, 25(12), 1264–77.
- Neely, A.D., Adams, C., and Kennerley, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Stakeholder Relationships*. London: Financial Times/Prentice Hall.
- Neely, A.D., Gregory, M.J., and Platts, K.W. (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations and Production Management*, 15(4), 80–116.
- Neely, A.D., Mills, J.F., Gregory, M.J., Richards, A.H., Platts, K.W., and Bourne, M.C.S. (1996). *Getting the Measure of Your Business*. Horton Kirby, Kent: Findlay Publications.
- Neely, A.D., Mills, J.F., Platts, K.W., Gregory, M.J., and Richards, A.H. (1994). Realising strategy through measurement. *International Journal of Operations and Production Management*, 14(3), 140–52.
- Neely, A.D., Richards, A.H., Mills, J.F., Platts, K.W., and Bourne, M.C.S. (1997). Designing performance measures: a structured approach. *International Journal of Operations and Production Management*, 17(11), 1131–53.
- Neely, A.D., and Wilson, J. (1992). Measuring product goal congruence: an exploratory study. *International Journal of Operations and Production Management*, 12(4), 45–52.
- New, C., and Szwajkowski, M. (1995). Performance measurement and the focussed factory: empirical evidence. *International Journal of Operations and Production Management*, 15(4), 63–79.
- O'Mahony, M., and Boer, W. (2002). Britain's relative productivity performance: has anything changed? *National Institute Economic Review*, 179, 38–43.
- Pande, P.S., Neuman, R.P., and Cavanagh, R. (2001). *The Six Sigma Way Team Fieldbook: An Implementation Guide for Project Improvement Teams*. New York: McGraw-Hill.
- Platts, K.W., and Gregory, M.J. (1990). Manufacturing audit in the process of strategy formulation. *International Journal of Operations and Production Management*, 10(9), 5–26.

- Plunkett, J.J., and Dale, B.G. (1988). Quality costs: a critique of some economic cost of quality models. *International Journal of Production Research*, 26(11), 1713–26.
- Porter, M., and Ketels, C. (2003). *UK Competitiveness: Moving to the Next Stage*. London: Department of Trade and Industry.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Ridgway, V.F. (1956). Dysfunctional consequences of performance measurements. *Administrative Science Quarterly*, 1(2), 240–7.
- Ruch, W.A. (1982). The measurement of white-collar productivity. *National Productivity Review*, 1(14), 416–26.
- Schonberger, R.J. (1982). *Japanese Manufacturing Techniques*. New York: Macmillan.
- (1986). *World Class Manufacturing*. New York: Macmillan.
- (1990). *Creating a Chain of Customers*. London: Guild Publishing.
- Sink, D.S. (1985). *Productivity Measurement: Planning, Measurement and Evaluation, Control and Improvement*. Chichester: John Wiley.
- Skinner, W. (1969). Manufacturing: missing link in corporate strategy. *Harvard Business Review*, 47(3), 136–45.
- (1974). The focused factory. *Harvard Business Review*, 52(3), 113–21.
- Slack, N. (1983). Flexibility as a manufacturing objective. *International Journal of Operations and Production Management*, 3(3), 4–13.
- (1987). The flexibility of manufacturing systems. *International Journal of Operations and Production Management*, 7(4), 35–45.
- (1991). *The Manufacturing Advantage: Achieving Competitive Manufacturing Operations*. London: Mercury.
- Slack, N., and Lewis, M. (2001). *Operations Strategy*. London: Prentice Hall.
- Stalk, G. (1988). Time: the next source of competitive advantage. *Harvard Business Review*, 66(4), 41–51.
- Turney, P.B.B., and Anderson, B. (1989). Accounting for continuous improvement. *Sloan Management Review*, 30(2), 37–48.
- Womack, J.P., Jones, D.T., and Roos, D. (1990). *The Machine that Changed the World*. New York: Rawson Associates.

4

Measuring performance: the supply chain management perspective

Douglas Lambert and Michael Knemeyer

Introduction

Increasingly, the management of relationships across the supply chain is being referred to as supply chain management (SCM).¹ In this environment, the ultimate success of the single company will depend on management's ability to integrate the company's intricate network of business relationships and more effectively manage these relationships (Drucker, 1998; Christopher, 1998). SCM offers the opportunity to capture the synergy of intra- and inter-company integration and management. In that sense, SCM deals with total business process excellence and represents a new way of managing the business and relationships with other members of the supply chain. The Global Supply Chain Forum (GSCF), a group of executives from non-competing organizations and a team of academic researchers, has been meeting regularly since 1992 with the objective of improving the theory and practice of SCM. The definition of SCM developed and used by the members of the GSCF follows (Lambert, 2006):

Supply Chain Management is the integration of key business processes from end user through original suppliers that provides products, services, and information that add value for customers and other stakeholders.

This broader view of SCM is illustrated in figure 4.1, which depicts a simplified supply chain network structure, the information and product flows, and the key supply chain management processes that penetrate the

The authors would like to acknowledge the contribution of the members of the Global Supply Chain Forum, whose practice, insight, ideas and comments have significantly contributed to this work. The Forum members include: 3M, Cargill, Coca-Cola Company, Colgate-Palmolive Company, Defense Logistics Agency, Hallmark, Hewlett-Packard Company, International Paper, Limited Brands, Masterfoods USA, Moen Incorporated, Shell Global Solutions International BV, TaylorMade-adidas Golf Company and Wendy's International.

¹ This chapter is adapted from Lambert (2006).

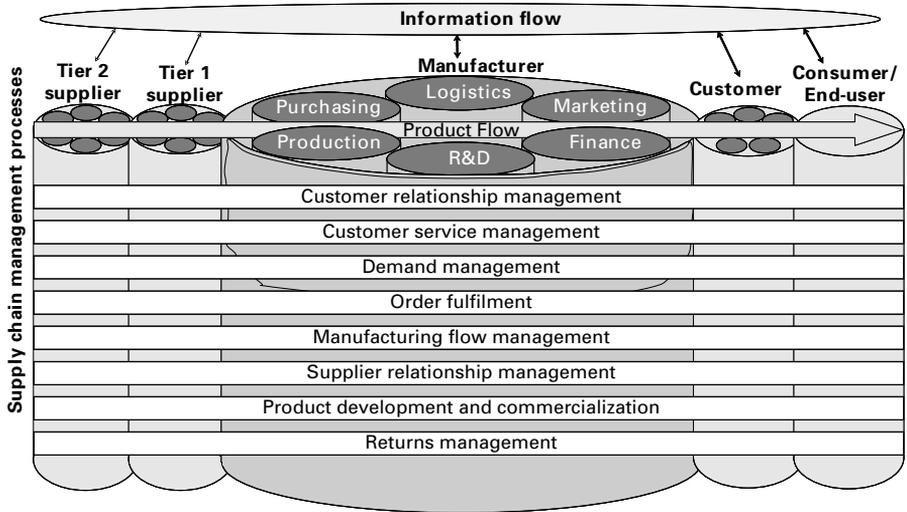


Figure 4.1: Supply chain management: integrating and managing business processes across the supply chain
Source: Lambert (2006, 3).

functional silos within the company and link with other entities in the supply chain. Thus, business processes become supply chain processes linked across intra- and inter-company boundaries. This process-based view of SCM goes beyond the perspective of SCM being a synonym for logistics, purchasing or operations, or a combination of these three functions. Successful SCM requires cross-functional integration of key business processes within the organization and across the network of organizations that comprise the supply chain. The challenge for management is to determine how to accomplish this integration successfully and develop performance measurement systems that support the achievement of this holistic perspective.

It is generally believed that a well-crafted system of supply chain metrics can increase the chances for success by aligning the SCM processes across multiple organizations, targeting the most profitable market segments and obtaining a competitive advantage through differentiated services and lower costs. The lack of proper metrics for a supply chain will result in failure to meet consumer/end-user expectations, suboptimization of company performance, missed opportunities to outperform the competition, and conflict within the organization and the supply chain. However, there is no evidence that meaningful performance measures that span the supply chain actually exist. Many factors may contribute to this situation, including a lack of supply chain or process orientation, the complexity of capturing metrics

across multiple organizations, little cooperation and sharing of information among organizations or the inability to capture performance information by customer, product or supply chain. A major contributor to the lack of meaningful supply chain performance measures is the absence of a framework for developing and designing such measures for the supply chain.

In most organizations, the metrics that management refers to as supply chain metrics are primarily internally focused logistics measures, such as lead time, fill rate or on-time performance. In many instances, these measures are financial (inventory turns and overall profitability) but they do not provide insight regarding how well key business processes have been performed or how effectively the supply chain has met customer needs. In a growing number of organizations management is beginning to measure performance outside the organization, but these efforts have largely been limited to evaluating the performance of tier 1 suppliers, customers or third-party providers (Lambert and Pohlen, 2001).

In this chapter we provide direction for those desiring to develop a holistic SCM performance measurement system. First, the conceptual framework of key SCM processes developed by the GSCF is presented. Second, we describe the problem with current metrics and establish the need for supply chain performance measures that take a more holistic viewpoint and are less likely to be “gamed” by members of the supply chain. Finally, the chapter concludes by establishing a framework for developing supply chain metrics. This framework identifies how each organization affects overall supply chain performance, and then suggests how this performance can be translated into shareholder value.

A conceptual framework for supply chain management

The GSCF conceptual framework emphasizes the interrelated nature of SCM and the need to proceed through several steps to design and successfully manage a supply chain. The SCM framework consists of three closely interrelated elements: the supply chain network structure, the supply chain business processes and the supply chain management components (see figure 4.2).

The supply chain network structure is comprised of the member organizations and the links between these organizations. Business processes are the activities that produce a specific output of value to the customer. The management components are the managerial variables by which the business

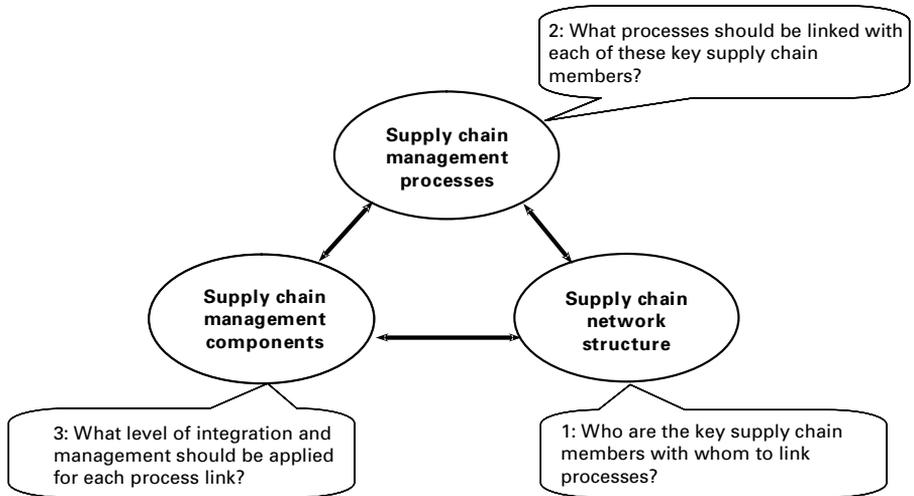


Figure 4.2: Supply chain management framework: elements and key decisions

Source: Lambert (2006, 9).

processes are integrated and managed across the supply chain. These three interrelated elements – supply chain network structure, supply chain management processes and supply chain management components – that together constitute the framework are described in the following sections.

Supply chain network structure

All organizations participate in supply chains, from the raw materials to the ultimate consumer. How much these supply chains need to be managed depends on several factors, such as the complexity of the product, the number of available suppliers and the availability of raw materials. Dimensions to consider include the length of the supply chain and the number of suppliers and customers at each tier. It would be rare for an organization to participate in only one supply chain. For most manufacturers, such as 3M, Colgate-Palmolive and Hewlett-Packard, the supply chain looks less like a pipeline or chain than an uprooted tree (see figure 4.3) where the branches and roots are the extensive network of customers and suppliers. The question is: how many of these branches and roots need to be managed?

Imagine the degree of complexity required to manage all suppliers back to the point of origin and all products/services out to the point of consumption. It is probably easier to understand why executives would want to manage

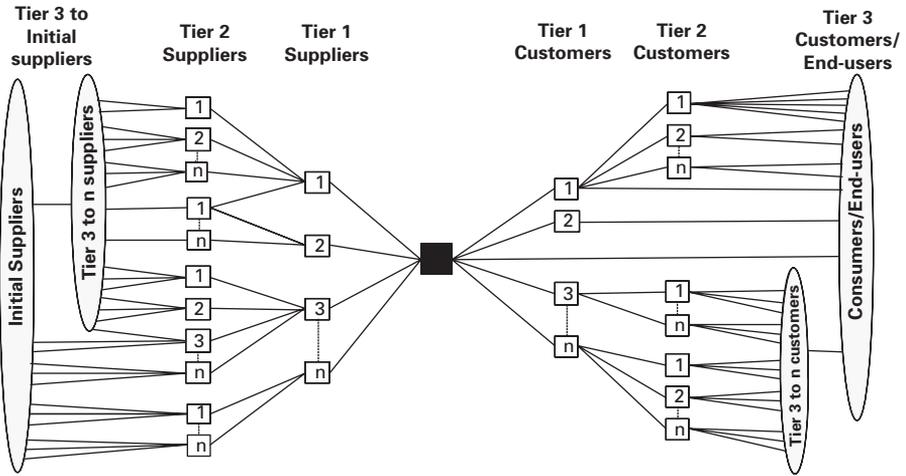


Figure 4.3: Supply chain network structure

Source: Lambert (2006, 5).

their supply chains to the point of consumption, because whoever has the relationship with the end-user has the power in the supply chain. Intel created a relationship with the end-user by having computer manufacturers place an “Intel inside” label on their computers. This affects the computer manufacturer’s ability to switch microprocessor suppliers. In any case, managing all tier 1 suppliers’ networks to the point of origin is an enormous undertaking, as illustrated in figure 4.3.

Management needs to choose the type of relationship appropriate for particular supply chain links (Lambert, Emmelhainz and Gardner, 1996). Not all links throughout the supply chain should be closely coordinated and integrated. The most appropriate relationship is the one that best fits the specific set of circumstances (Cooper and Gardner, 1993). Determining which parts of the supply chain deserve management attention must be weighed against the organization’s capabilities and the importance to the organization.

Supply chain management processes

Successful supply chain management requires a change from managing individual functions to integrating activities into supply chain processes. In many major corporations, such as 3M and Coca-Cola, management has

reached the conclusion that optimizing the product flows cannot be accomplished without implementing a process approach. The key supply chain processes identified by members of the GSCF are as follows:

- customer relationship management;
- customer service management;
- demand management;
- order fulfilment;
- manufacturing flow management;
- supplier relationship management;
- product development and commercialization; and
- returns management.

These processes are shown in figure 4.1. A description of each of the eight processes follows (Lambert, 2006; Croxton *et al.*, 2001).

Customer relationship management

CRM provides the structure for how relationships with customers are developed and maintained. Management identifies key customers and customer groups to be targeted as part of the organization's business mission. The goal is to segment customers based on their value over time and increase customer loyalty by providing customized products and services. Cross-functional customer teams tailor product and service agreements (PSAs) to meet the needs of key accounts and for segments of other customers. The PSAs specify levels of performance. The teams work with key customers to improve processes and eliminate demand variability and non-value-added activities. Performance reports are designed to measure the profitability of individual customers as well as the organization's financial impact on the customer.

Customer service management

The customer service management process is the organization's face to the customer. It provides the key point of contact for administering the PSA. The customer service management process involves monitoring customer PSAs, and identifying and solving problems before the customer is affected. Customer service provides the customer with real-time information on shipping dates and product availability through interfaces with the organization's functions and the other processes.

Demand management

Demand management is the SCM process that balances the customers' requirements with the capabilities of the supply chain. With the right process in place, management can match supply with demand proactively and execute the plan with minimal disruptions. In particular, if managers are proactively managing the demand management process, they need to manage their company's activities that influence customer demand patterns, such as end-of-quarter promotions or financial terms of sale that cause customers to buy in strange ways. The process is not, therefore, limited to forecasting. It includes synchronizing supply and demand, increasing flexibility and reducing variability. A good demand management system uses point-of-sale and key customer data to reduce uncertainty and provide efficient flows throughout the supply chain. In advanced applications, customer demand and production rates are synchronized to manage inventories globally.

Order fulfilment

The order fulfilment process involves more than just filling orders. It includes all activities necessary to define customer requirements, design a logistics network and enable an organization to meet customer requests while minimizing the total delivered cost as well as filling customer orders. This is not just the logistics function but, instead, needs to be implemented cross-functionally and with the coordination of key suppliers and customers. For example, in complex global organizations, the finance function provides important information regarding tax rates, tariffs and exchange rates that will influence how various network configurations affect overall profitability. The objective is to develop a seamless process from the supplier to the organization and to its various customer segments.

Manufacturing flow management

Manufacturing flow management is the SCM process that includes all activities necessary to move products through the plants and to obtain, implement and manage manufacturing flexibility in the supply chain. Manufacturing flexibility reflects the ability to make a wide variety of products in a timely manner at the lowest possible cost. To achieve the desired level of manufacturing flexibility, planning and execution must extend

beyond the four walls of the manufacturer and to other members of the supply chain.

Supplier relationship management

Supplier relationship management (SRM) is the process that defines how a company interacts with its suppliers. As the name suggests, this is a mirror image of customer relationship management. Just as a company needs to develop relationships with its customers, it also needs to foster relationships with its suppliers. As in the case of CRM, a company will forge close relationships with a small subset of its suppliers and manage arm's-length relationships with others. A PSA is negotiated with each key supplier, defining the terms of the relationship. For segments of less critical suppliers, the PSA is not negotiable. SRM is about defining and managing these PSAs. Long-term partnerships are developed with a small group of core suppliers. The desired outcome is a win-win relationship whereby both parties benefit.

Product development and commercialization

Product development and commercialization constitute the SCM process that provides the structure for developing and bringing products to market jointly with customers and suppliers. The product development and commercialization process team must coordinate with the CRM process teams to identify customers' articulated and unarticulated needs; select materials and suppliers in conjunction with the SRM process teams; and develop production technology with the manufacturing flow management process team to manufacture and provide the best supply chain flow for the product/market combination.

Returns management

Returns management is the SCM process by which activities associated with returns, reverse logistics, gatekeeping and avoidance are managed within the organization and across key members of the supply chain. The correct implementation of this process enables management not only to manage the reverse product flow efficiently but also to identify opportunities to reduce unwanted returns and to control reusable assets, such as containers. The concept of returns avoidance is a key aspect of this process that differentiates it from reverse logistics. For example, the catalogue business at Victoria's Secret, a lingerie retailer, incurs many returns as a result of sizing issues. In order to

reduce the number of returns, management works with suppliers to apply sizing guidelines across all products in a uniform manner. This reduces the costs associated with returns and improves customer satisfaction. Effective returns management is an important part of SCM and provides an opportunity to achieve a sustainable competitive advantage.

The management components of SCM

The management components of SCM are the third element of the SCM framework (see figure 4.2). The level of integration and management of a business process link is a function of the number and level, ranging from low to high, of components added to the link (Lambert, Emmelhainz and Gardner, 1996). Consequently, adding more management components or increasing the level of each component can increase the level of integration of the business process link.

The literature on business process re-engineering (Hammer and Champy, 1993; Towers, 1994; Hewitt, 1994), buyer–supplier relationships (Ellram and Cooper, 1990; Stevens, 1989) and SCM (Cooper, Lambert and Pagh, 1997; Olsen and Ellram, 1997) suggests numerous possible components that must receive managerial attention when managing supply chain relationships. Based on the management components identified in our previous work, review of the literature and interviews with eighty managers, nine management components are identified for successful SCM:

- planning and control;
- work structure;
- organization structure;
- product flow facility structure;
- information flow;
- management methods;
- power and leadership structure;
- risk and reward structure; and
- culture and attitude.

Planning and control of operations are keys to moving an organization or supply chain in a desired direction. The extent of joint planning is expected to bear heavily on the success of the supply chain; different components may be emphasized at different times during the life of the supply chain, but planning transcends the phases (Cooper *et al.*, 1997). The control aspects can be operationalized as the best performance metrics for measuring supply chain success.

The work structure indicates how the organization performs its tasks and activities. The level of integration of processes across the supply chain is a measure of organizational structure. All but one of the literature sources examined cited work structure as an important component. Organizational structure can refer to the individual organization and the supply chain; the use of cross-functional teams would suggest more of a process approach. When these teams cross organizational boundaries, such as in-plant supplier personnel, the supply chain should be more integrated.

Product structure refers to the flow of product through the network for sourcing, manufacturing and distribution across the supply chain. Since inventory is necessary in the system, some supply chain members may keep a disproportionate amount of inventory. As it is less expensive to have unfinished or semi-finished goods in inventory than finished goods, upstream members may bear more of this burden. Rationalizing the supply chain network has implications for the performance of all members.

Virtually every author indicates that the information flow structure is important. The kind of information passed among channel members and the frequency of information updating has a strong influence on the efficiency of the supply chain. This may well be the first component integrated across part or all of the supply chain.

Management methods include the corporate philosophy and management techniques. It is very difficult to integrate a top-down organization structure with a bottom-up structure. The level of management involvement in day-to-day operations can differ across supply chain members.

The power and leadership structure across the supply chain will affect its form. One strong leader will drive the direction of the chain. In most supply chains studied to date, there are one or two strong leaders among the organizations. The exercise of power, or lack thereof, can affect the level of commitment of other members. Forced participation will encourage exit behaviour, given the opportunity (Macneil, 1980; Williamson, 1975). The anticipation of sharing risks and rewards across the supply chain affects the long-term commitment of its members.

The importance of corporate culture and its compatibility across members of the supply chain should not be underestimated. Meshing cultures and individuals' attitudes is time-consuming, but it is necessary at some level for the channel to perform as an integrated chain. Aspects of culture include how employees are valued and incorporated into the management of the organization.

The problems with existing metrics

The performance measures used in most organizations have several problems that prevent them from effectively measuring supply chain performance. Many measures identified as “supply chain metrics” are actually measures of internal logistics operations rather than measures of supply chain management. The majority are single organizational logistics functional silo measures, such as fill rate, lead time, on-time performance, damage and responsiveness (Gilmour, 1999), and are not the multi-organization measures that are necessary to measure the performance of the supply chain (Beamon, 1989; Keebler *et al.*, 1999). Similar results were obtained in seminar programmes held at multiple locations, in the United States and abroad, when executives were asked to identify examples of supply chain metrics. Typically, the executives identified inventory turns as one of the measures of supply chain performance – a view shared by several authors (Anderson, Franke and Donavon, 1997; Lapide, 2000; Deloitte Consulting, 1999; Supply Chain Council, 2000; Cap Gemini Ernst and Young/Industry Week, 2000; Langdon, 2001). As a supply chain metric, however, an inventory turn is not an effective measure, providing a useful example of why new metrics are needed for managing the supply chain.

An inventory turn rate fails to capture key differences in product cost, form and risk within the supply chain. Figure 4.4 illustrates inventory positions and flows across a supply chain and helps make this point. As inventory moves closer to the point of consumption, it increases in value – that is, the out-of-pocket cash investment in the inventory increases. Consequently, if the opportunity cost of money and the inventory turns are similar, the inventory carrying cost (ICC) is much higher at the retail level, and an inventory turn improvement by the retailer has a much greater effect on overall supply chain performance than a turn improvement by the supplier, or manufacturer, and a greater impact than a turn improvement by the wholesaler (see table 4.1).

Referring to table 4.1, if the supplier, manufacturer, wholesaler and retailer are all achieving six turns, an improvement to seven turns would be worth \$0.04, \$0.21, \$0.53 and \$0.62 per unit sold, respectively. This example demonstrates that the common practice of pushing inventory forward in the supply chain reduces overall supply chain performance unless increased sales offset the associated increased costs. However, improvements in inventory performance must also reflect current inventory turns at each level within the

Table 4.1: How supply chain position affects inventory carrying cost

		Supplier	Manufacturer	Distributor/ Wholesaler	Retailer
Cash value of inventory*		\$5	\$25	\$62	\$72
Inventory carrying cost(%)		36	36	36	36
ICC/unit with:	1 turn	\$1.80	\$9.00	\$22.32	\$25.92
	2 turns	\$0.90	\$4.50	\$11.16	\$12.96
	3 turns	\$0.60	\$3.00	\$7.44	\$8.64
	4 turns	\$0.45	\$2.25	\$5.58	\$6.48
	5 turns	\$0.36	\$1.80	\$4.46	\$5.18
	6 turns	\$0.30	\$1.50	\$3.72	\$4.32
	7 turns	\$0.26	\$1.29	\$3.19	\$3.70
	8 turns	\$0.23	\$1.13	\$2.79	\$3.24
	9 turns	\$0.20	\$1.00	\$2.48	\$2.88
	10 turns	\$0.18	\$0.90	\$2.23	\$2.59
	11 turns	\$0.16	\$0.82	\$2.03	\$2.36
	12 turns	\$0.15	\$0.75	\$1.86	\$2.16

Source: Lambert (2006, 200).

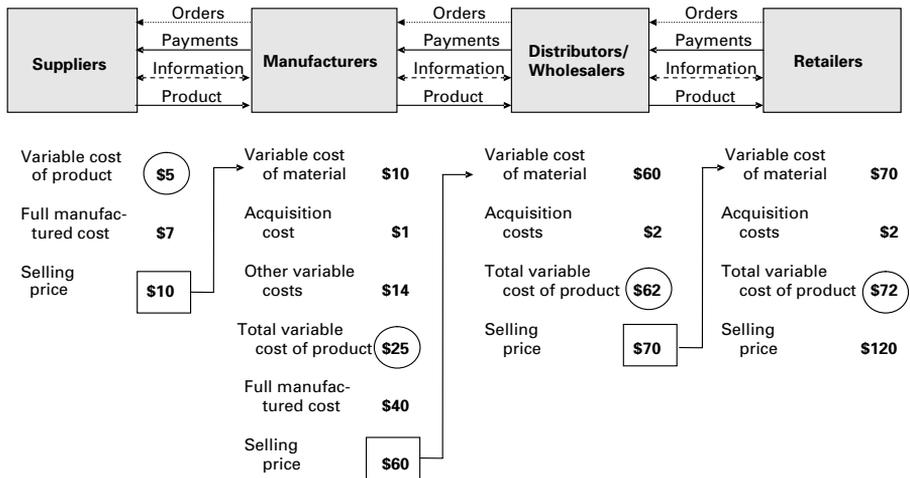


Figure 4.4: Inventory flows within the supply chain

Source: Lambert (2006, 199).

supply chain. Figure 4.5 shows graphically how the inventory carrying cost per unit changes with the number of inventory turns. While figure 4.5 reflects the manufacturer data from table 4.1, the shape of the curve is identical for the supplier, wholesaler/distributor and retailer.

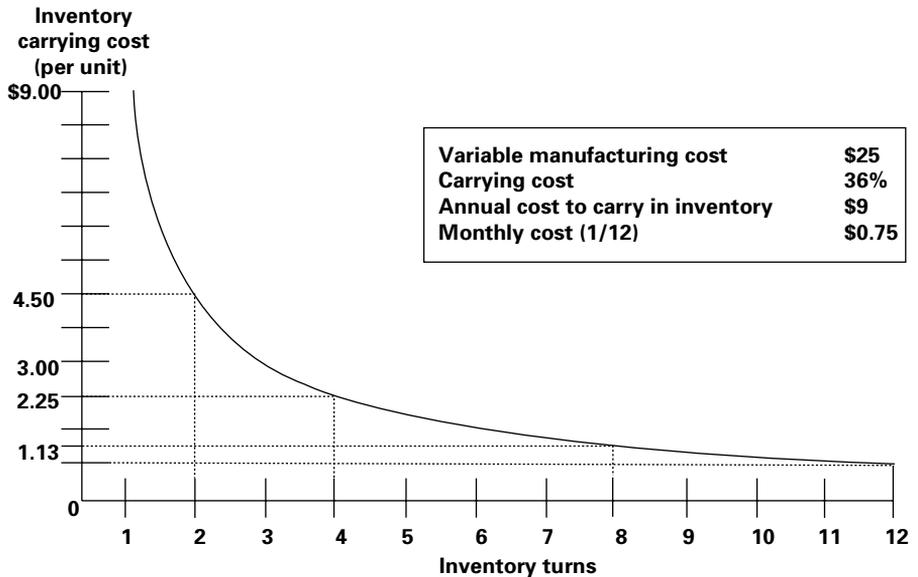


Figure 4.5: Annual inventory carrying costs compared to inventory turns for the manufacturer in table 4.1

Source: Lambert (2006, 200).

Now let's reconsider how existing inventory turns at various tiers in the supply chain will affect the general rule that inventory, or inventory ownership, should be moved backward in the supply chain. If the manufacturer is achieving only five turns and the wholesaler has eleven turns, a one-turn improvement equates to \$0.30 per unit sold for the manufacturer and \$0.17 per unit sold for the wholesaler. In this case, the general rule is broken.

A single inventory turn metric for the supply chain cannot capture the differences that an improvement in turns will have at each level or for the total supply chain. Performance, as measured by total inventory carrying costs, would be a better measure since it considers both the cash value of the inventory at various positions in the supply chain as well as varying opportunity costs for inventory investments (Stock and Lambert, 2001). Total ICC is improved by "pushing" inventory backwards in the supply chain towards the point of origin. The further back inventory is pushed the lower the overall inventory carrying costs for the entire supply chain.

In addition, an inventory turn rate does not recognize the different forms or the risk of the inventory. Raw materials held by the supplier may be used for multiple purposes, products or customers. The further downstream the inventory the greater the risk that it does not exactly meet consumers'

requirements. Pushing the inventory backwards and postponing its final form permits the supply chain to avoid higher obsolescence and repositioning costs. Additionally, if it is not dedicated inventory, this approach should leverage an aggregation effect across customers that would enable the supply chain to carry a smaller amount of inventory. In summary, inventory turns and other commonly used logistics measures are inadequate for evaluating and aligning performance across multiple organizations in the supply chain (García-Dastugue and Lambert, 2007).

Another problem with metrics stems from the lack of a widely accepted definition for supply chain management. Until recently, many logistics practitioners, academics and consultants viewed SCM as an extension of logistics outside the organization to include customers and suppliers (Coyle, Bardi and Novack, 2000; Copacino, 1997; Simchi-Levi, Kaminsky and Simchi-Levi, 2000). As previously described, successful SCM requires cross-functional integration of key business processes within the organization and across the network of organizations that comprise the supply chain.

Why supply chain metrics?

Several factors contribute to management's need for new types of measures for managing the supply chain.

- The lack of measures that capture performance across the entire supply chain.
- The requirement to go beyond internal metrics and take a supply chain perspective.
- The need to determine the relationship between individual company performance and overall supply chain performance.
- The complexity of supply chain management.
- The requirement to align activities and share joint performance measurement information to implement strategy that achieves supply chain objectives.
- The desire to expand the “line of sight” within the supply chain.
- The requirement to allocate benefits and burdens resulting from functional shifts within the supply chain.
- The need to differentiate the supply chain to obtain a competitive advantage.
- The goal of encouraging cooperative behavior across corporate functions and across organizations in the supply chain.

Measures spanning the entire supply chain do not exist (Mentzer, 2001), and existing logistics and other functional measures have not kept up with the changing role and scope of supply chain management (Caplice and Sheffi, 1995). As a result, managers are able to determine whether they have met their corporate goals only after the fact, by diagnosing poor financial results or when they lose a key customer (Lapide, 2000). Moreover, measures have little to do with supply chain strategy and objectives and may actually conflict, resulting in inefficiencies for the overall supply chain (Lee and Billington, 1992). Metrics integrating performance across multiple organizations are just emerging (Francella and Doherty, 1998), but they are only in their early stages and have generally focused on measuring the performance of adjacent channel members: suppliers, carriers and immediate customers.

An area of research that deals with performance measurement across the supply chain is “time compression” (Wilding, 2002; Christopher, 2005). Time, especially an analysis of “value-added” versus “non-value-added” time, has the ability to provide management with improvement opportunities. However, it should be pointed out that, like any other “single” measure of performance, it has the potential to be “gamed”. When time is needlessly long, reductions may improve performance for all members of the supply chain. At some point, however, there is the potential that further reductions in time could result in lost sales and higher costs. Although time provides information that management can use to collaborate and improve performance, it falls short of assessing how the competitiveness, customer value and shareholder value for each of the respective organizations in the supply chain are affected. One measure of time is a day of inventory. As in the example of inventory turns, the financial impact of a day of inventory at various tiers of the supply chain is not the same.

The adoption of a supply chain approach holds numerous consequences for the measurement and control of individual business activities (van Hoek, 1998) and the performance measures used. The shift from a functional to a process focus will require the development of new types of measures, financial as well as information about the operational performance of these processes (Kallio *et al.*, 2000). Supply chains, rather than the functional operations within a single company, will become the new focus (Keebler *et al.*, 1999). Supply chain members will become responsible and accountable for the joint performance measures of these key business processes, and they will require an integrated information system to enable multiple members of the supply chain to gain access to these measurers (Lee, 2000). Managers will need to understand not only their own activities and

costs but those of their organization's upstream and downstream supply chain members (Francella and Doherty, 1998).

Linking corporate and supply chain performance

In order for management to understand the interrelationship between corporate and supply chain performance, more holistic measures are required to quantify projected changes in supply chain performance and relate them to changes in corporate financial and non-financial performance (ECR Performance Measures Operating Committee, 1994). The translation of these measures into corporate financial performance and shareholder value will be critical for resolving conflicting objectives and supporting cost trade-offs across the supply chain, especially in areas where cost or asset increases will be required. Existing measurement systems provide little assistance or insight regarding the question "What's in it for me?" (van Hoek, 1998). Future SCM innovations will come under increasing scrutiny to determine if and when they yield a positive impact on corporate performance.

The complexity of the supply chain requires a different approach for designing metrics and measuring performance. As mentioned previously, a supply chain can be represented as an uprooted tree, where the roots are the suppliers and the branches are the customers. The complexity of most supply chains makes it difficult to understand how activities at multiple tiers are related and influence each other. Performance measures must reflect this complexity and consider cross-company operations from suppliers all the way to the customer (Francella and Doherty, 1998). Such an approach enables members of the supply chain to alter their efforts to serve the end-consumer better.

Relationship between supply chain metrics and strategy

Implementing a supply chain strategy requires metrics that align performance with the objectives of other members of the supply chain (Francella and Doherty, 1998). Managers can no longer focus solely on optimizing their own organization's operations. They need to work collaboratively to generate the greatest mutual gains and savings (Keebler *et al.*, 1999). Aligned metrics can assist in shifting managers' focus to attaining the operational goals of the supply chain (Walker, 1999). The

alignment of metrics enables managers to identify and institutionalize the organizational, operational and behavioural changes (Mercer Management Consulting, undated) needed to manage the key business processes spanning their supply chain network. However, integrating the key business processes across the supply chain is difficult because of the many constituencies, each with their own metrics and individual objectives (Sherman, 1998). These objectives may have little in common, resulting in potential conflict and inefficiencies for the supply chain (Lee and Billington, 1992). Conflicting objectives will preclude managers from effectively managing trade-offs across functions (Lee and Billington, 1992) as well as across organizations.

Managers need to extend their “line of sight” across the supply chain by measuring the performance of activities and organizations they do not directly control (Lapide, 2000; Reese, 2001). A single company rarely controls the entire supply chain and cannot “see” supply-chain-wide areas for improvement (van Hoek, 1998). Increased visibility and shared metrics will assist management with the integration and synchronization of inter-enterprise processes. The visibility makes the supply chain more transparent and can lead to performance improvements. Managers can determine how well the supply chain performed against the expectations of their customers (Reese, 2001) and use the information to determine where performance improvements need to occur. The identification of deficiencies outside a company’s span of control can lead to programmes aimed at improving performance, or taking some level of control of upstream or downstream supply chain activities.

Functional shifts and cost trade-offs made across multiple organizations require metrics that management can use to measure and allocate the resulting benefits and burdens. Individual organizations may have to sacrifice internal efficiencies or perform additional functions to reduce or “optimize” total supply chain costs (van Hoek, 1998). Consequently, some organizations will benefit from the realignment of activities of functions while others will incur additional “burdens” or costs. Management needs the capability to measure where any benefits or burdens have occurred and to have a mechanism for equitably redistributing the benefits across the involved members of the supply chain (La Londe and Pohlen, 1994).

Supply chain metrics are needed to sustain competitiveness and to differentiate product and service offerings. The commoditization of products is forcing management to examine the supply chain to determine opportunities to increase revenues and eliminate costs (Keebler *et al.*, 1999). In addition, there are more options for the organization to be distinguished

from competitors in terms of supply chain capabilities. Integrated metrics will allow management to assess the competitiveness of the supply chain as a whole and to determine which internal improvement efforts will produce the greatest impact on overall competitiveness (van Hoek, 1998).

Supply chain metrics are also required to encourage the desired behaviour and attain the level of performance required in the supply chain. Rewards and incentives are usually based on performance measurements that are focused internally rather than on the consumer or the supply chain (Neely *et al.*, 1995). Measurement is important, as it influences the behaviour of individuals who determine supply chain performance (Lapide, 2000). Management needs metrics to determine whether the performance of the organization's supply chain members has improved or degraded and what factors have contributed to the situation.

Requirement to align performance measures across the supply chain

Many of the measurements used within organizations are developed internally, in isolation, and are linked to local rewards rather than strategy. The missing connection between strategy and measurements promotes an internal corporate focus that becomes an obstacle to developing supply chain measurement systems (Mentzer, 2001; Kuwaiti and Kay, 2000). The disconnect between strategy and performance measures further contributes to many of the strategy-level measures appearing unrelated or not actionable at lower levels in the corporate hierarchy. For example, many critical customer service measures are frequently not tracked by existing performance measurement systems despite their effect on supply chain performance (Lee and Billington, 1992). As a result, the relationship between the performance of supply chain activities and what creates value for the consumer is not clearly understood (Francella and Doherty, 1998).

Shifting the orientation of performance measurement from an internal corporate view to that of the supply chain provides several benefits. Managers will obtain a better understanding of how their performance creates consumer value and are more likely to know where bottlenecks or problems exist in their supply chains (Keebler *et al.*, 1999). The exchange of performance information will encourage closer coordination and collaboration (Lapide, 2000; Holmberg, 2000; Lee, 2000). By focusing on the entire supply chain the metrics encourage innovative solutions, since managers will explore those alternatives with the possibility of meeting the supply

chain's performance objectives (Caplice and Sheffi, 1995). Other benefits include: significant reductions in operating costs, substantial improvements in customer service, the pursuit of new growth opportunities, faster response times and reduced inventory investments (Keebler *et al.*, 1999).

In a small but growing number of organizations, management has begun to expand metrics to include the performance of customers and/or suppliers (Lee, 2000; Levy *et al.*, 1995). The inclusion of supplier and customer performance represents a step in the right direction, but it still does not provide a means for measuring supply chain performance. These performance measures continue to focus on measuring logistics performance but now across several organizations in the supply chain (Keebler *et al.*, 1999; Levy *et al.*, 1995). A focus limited to logistics does not encompass the key business processes or the importance of other key corporate functions, thereby ignoring their potential effects on overall supply chain performance (Cooper *et al.*, 1997; Lambert and Cooper, 2000).

A framework for developing supply chain metrics

Complexity makes the development of supply chain performance metrics extremely difficult (see figure 4.6). For example, consumer goods manufacturers such as Colgate-Palmolive, Procter and Gamble, and Unilever in most cases sell to exactly the same customers, and in many cases purchase from the same suppliers. Competing supply chains appear more like interconnected or overlapping networks rather than as a mutually exclusive "supply chain versus supply chain" form of competition. The overlap results in many instances of shared inventories, shared services and shared assets among supply chains (Rice and Hoppe, 2001). As a result, it is not easy for managers to determine how business practices with specific organizations drive total supply chain performance. As we pointed out earlier, you cannot simply add up inventory turns across organizations in the supply chain.

Despite the complexity and overlap existing in most supply chains, managers can develop metrics to align the performance of key business processes across multiple organizations. We propose a framework that aligns performance at each link (supplier–customer pair) within the supply chain. The framework begins with the linkages at the focal company and then moves outward a link at a time. The link-by-link approach provides a means for aligning performance from the point of origin to the point of consumption, with the overall objective of maximizing shareholder value for the total supply

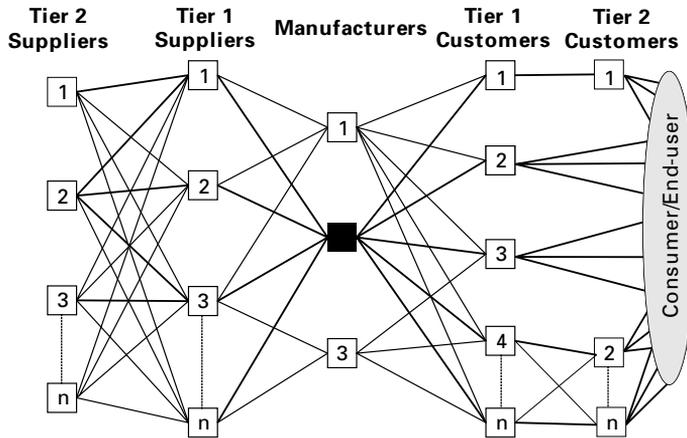


Figure 4.6: Supply chain complexity

Source: Lambert (2006, 208).

chain, as well as increasing shareholder value for each company in the supply chain. When managers hold the viewpoint that their supply chain must be managed link by link and relationship by relationship, competitive advantage is achieved by managing those linkages more effectively than competitors. In addition, the use of profit and loss (P&L) statements as prescribed by this framework helps to ensure that “gaming” of performance measures does not occur. The framework consists of seven steps.

- Map the supply chain from the point of origin to the point of consumption to identify where key linkages exist in the supply chain.
- Use the customer relationship management and supplier relationship management processes to analyse each link (customer–supplier pair) and determine where additional value could be created for the supply chain.
- Develop customer and supplier P&L statements to assess the affect of the relationship on profitability and shareholder value.
- Realign supply chain processes and activities to achieve performance objectives.
- Establish operational performance measures that align individual behaviour with supply chain process objectives, profitability and shareholder value.
- Compare resulting shareholder value and market capitalization across organizations with objectives; revise process and performance measures as necessary.
- Replicate steps at each link in the supply chain.

Map the supply chain

The framework begins with the mapping of the supply chain from the point of origin to the point of consumption. The map identifies the various paths that materials and information flows may take from source to the final consumer (see figure 4.3). Managers can use the map to identify the key supply chain linkages that are most critical to success. The initial focus should be on managing those dyads of the supply chains with the greatest potential for increasing profitability and developing a sustainable competitive advantage. CRM and SRM are the two major processes that connect organizations and can be used to link up the entire supply chain.

Analyse each link

The supplier applies the CRM process to define how it will manage relationships with customers. Key customers are identified and the supplier's CRM teams work with these accounts to tailor product and service agreements that meet their requirements and specify the level of performance. The CRM process creates value by working with the customer to improve performance (see figure 4.7). For example, the CRM team may negotiate with the customer's team to implement supplier-managed inventory (SMI). Successful SMI implementation may lead to increased revenues as the customer allocates a larger proportion of the business to that supplier. If the relationship reduces costs and can yield a price reduction for the consumer, revenues may increase as total sales for the supply chain increase. The cost of goods sold may decrease through better scheduling of material requirements and utilization of plant capacity and labour. The supplier's expenses may increase as the company assumes ownership of, and responsibility for, the customer's inventory; however, other expenses may decrease due to reduced order processing and forecasting costs. Inventory carrying costs will decrease as point-of-sale data are used to schedule shipments instead of forecasting requirements and maintaining safety stock. Better capacity utilization and collaborative planning and forecasting of requirements may reduce the need for customer-specific assets. The process improvements obtained through CRM can be translated into increased shareholder value through the use of an economic value added model, as illustrated in figure 4.7.

On the opposite side of the dyad, the customer uses the SRM process to manage supplier relationships. The customer selects and develops relationships with suppliers based on their criticality. As with the CRM process, it is

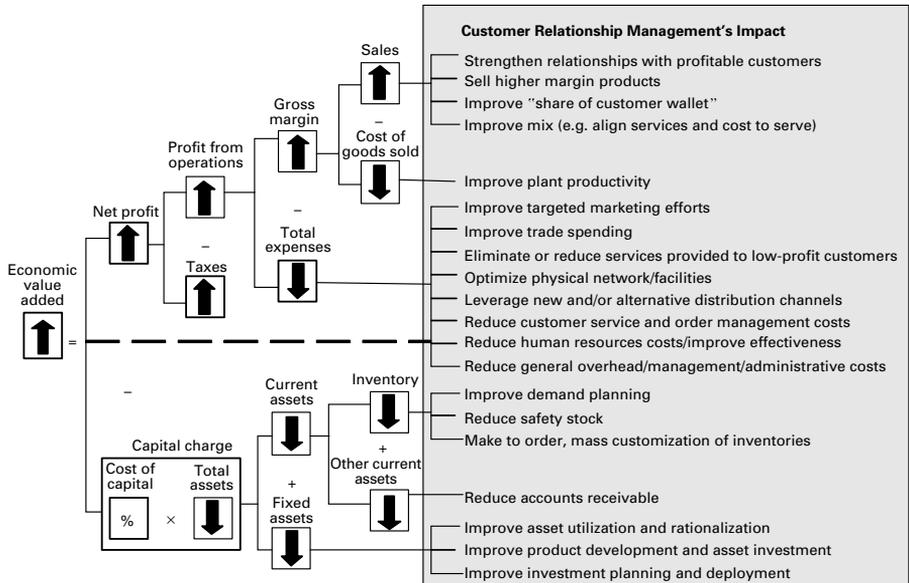


Figure 4.7: How customer relationship management affects economic value added

Source: Lambert (2006, 210).

possible to identify how SRM affects EVA[®] (see figure 4.8). Referring to the previous example, the SRM process would capture the value created through SMI implementation. The relationship may produce increased revenues through cost reduction, lower consumer prices and improved quality obtained by working with a select group of suppliers. The cost of goods sold (COGS) may be reduced through the leveraging of larger buys with a smaller number of suppliers. Expenses decrease as the supplier assumes responsibility for order placement and inventory management. Pushing the ownership of inventory backwards to the supplier reduces ICCs for the customer and for the total supply chain, since the supplier owns the inventory at a lower cash value (see figure 4.4). Together, the SRM and CRM processes capture the total value, adjusted for the cost of money, created by the supplier–customer relationship. Similar charts can be developed for the other six processes.

Develop profit and loss statements

The development of customer and supplier P&L statements provides a complete picture of how the relationship affects profitability for both organizations (see figure 4.9). Initiatives undertaken by the two organizations will be

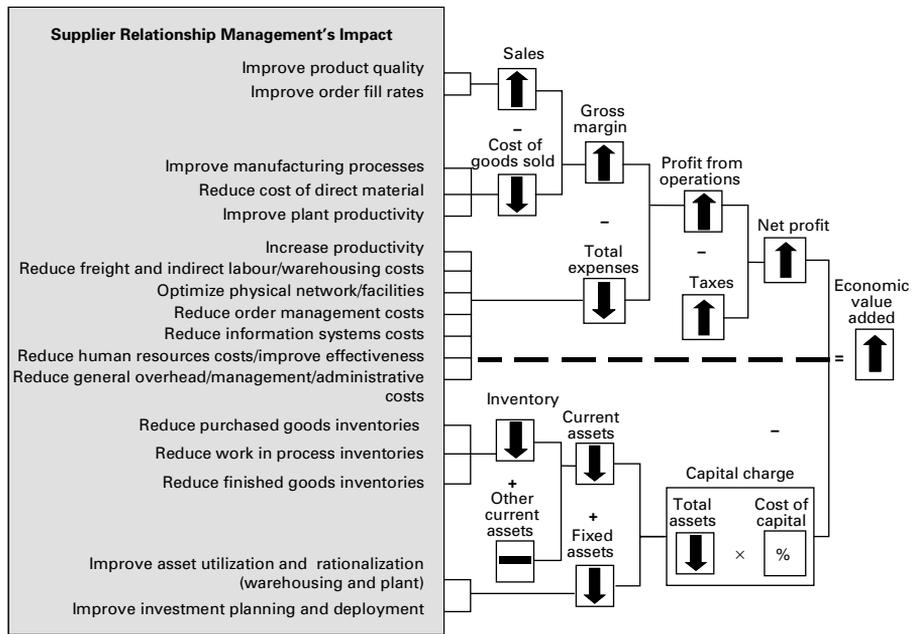


Figure 4.8: How supplier relationship management affects economic value added

Source: Lambert (2006, 210).

reflected in these P&Ls, as will improvements in performance of the other six processes (see figure 4.1). While performance metrics must be developed for all eight processes in order to motivate the desired behaviour, the financial performance of all eight processes is captured in the customer P&Ls. When the customer P&Ls are aggregated for all customers and corporate joint costs are deducted, the results represent overall organization performance.

Figure 4.9 represents a manufacturer selling to a wholesaler/retailer. In the case of the supplier (manufacturer), variable manufacturing costs are deducted from net sales to calculate a manufacturing contribution. Next, variable marketing and logistics costs are deducted to calculate a contribution margin. Assignable non-variable costs, such as slotting allowances and ICCs, are subtracted to obtain a segment controllable margin. The net margin is obtained after deducting a charge for dedicated assets. In the case of the customer (wholesaler/retailer), product costs are deducted from sales to obtain a gross margin, to which discounts and allowances are added to obtain the net margin. The remaining steps are similar to the approach taken for the supplier to obtain the net segment margin. These statements contain opportunity costs for investments in receivables, inventory and other assets

Supplier	CUSTOMER		SUPPLIER A
	A	Customer	
Net sales		Sales	
Cost of goods sold (var. mfg. cost)		Cost of goods sold	
Manufacturing contribution		Gross margin	
		Plus: discounts and allowances	
Variable marketing and logistics costs:		Market development funds	
Sales commissions		Slotting allowances	
Transportation		Co-op advertising	
Warehousing (Handling in and out)		Net margin	
Special packaging		Variable marketing and logistics costs:	
Order processing		Transportation	
Charge for investment in accts. rec.		Receiving	
Contribution margin		Order processing	
		Contribution margin	
Assignable non-variable costs:		Assignable non-variable costs:	
Salaries		Salaries	
Segment-related advertising		Advertising	
Slotting allowances		Inventory carrying costs less:	
Inventory carrying costs		Charge for accounts payable	
Segment controllable margin		Segment controllable margin	
		Charge for dedicated assets used	
Charge for dedicated assets used		Net segment margin	
Net segment margin			

Figure 4.9: Combined customer–supplier profitability analysis

Source: Lambert (2006, 210).

dedicated to the relationship. Consequently, they are much closer to cash flow statements than a traditional P&L. If the supplier is selling an undifferentiated commodity to a customer that is another manufacturer, then the customer’s report on the supplier would be a total cost analysis. In this case, the change in the total cost would represent the change in shareholder value, just as the change in the profit as measured by customer P&Ls would measure the change in shareholder value.

Realign supply chain management processes

The P&Ls provide the ultimate measure of supply chain performance and can be used to align processes and organizations. In our SMI example, the supply chain has the objective of increasing availability while simultaneously reducing costs. SMI implementation may cause the supplier to incur additional costs in some areas while obtaining cost reductions in others. The supplier’s P&L will reflect the resulting total cost as well as changes in assets (because of charges for assets employed), revenue and profitability. The customer’s P&L will, similarly, reflect any changes due to SMI implementation. A combined profitability analysis captures the total effort and enables managers to reach a better understanding of how aligning their action with supply chain objectives drives profitability in their organizations. They can then use this information as a basis for negotiating how to split equitably any benefits or burdens resulting from supply chain process

improvements. This approach provides a much more accurate measure of supply chain performance than existing measures.

Functional or logistics measures, such as inventory turns, cannot capture the full extent of management cost trade-offs and can easily be “gamed”. As pointed out earlier in this chapter, inventory carrying cost is a better measure, but it does not capture the costs incurred to achieve the reduction in inventory. The increase in transportation and ordering costs may more than offset any gains made in ICCs. As described earlier, inventory reductions typically have a greater impact on total supply chain performance if they occur at the retail level. Generally speaking, making to order and pushing inventory backwards in the supply chain improves overall performance. A combined supplier–customer profitability analysis will capture how the repositioning of inventory improves total supply chain performance, whereas inventory turns does not reflect any of the cost trade-offs either within an organization or in the supplier–customer link.

Align operational measures with P&Ls

The P&Ls and EVA[®] measures alone are insufficient to effect improvements in supply chain performance and to align behaviour. Supply chain and corporate metrics must be cascaded down to develop performance measures at the levels in the organization where the work is done. For example, in the case of the order fulfilment process, managers would begin with the high-level objectives of organizational performance identified in figure 4.10, but it is also necessary to develop more specific measures of performance that connect to these high-level drivers. For example, at the operational level, warehouse employees may be given order picking accuracy and order picking productivity goals. While improvements in picking accuracy and productivity may be adequate for evaluating warehouse employees, those who set improvement goals ought to measure how these improvements affect EVA[®].

Warehouse employees supporting this process may not be able to relate how more efficient order picking impacts profitability or shareholder value, but they can focus on reducing order pick time and errors. Reducing order pick time while increasing productivity also reduces the cost per order. Reducing order pick errors will result in faster payment of customer invoices and reduces the cost of returned goods. By outperforming the competition, a faster order cycle time may positively affect sales, either through a larger market share or more repeat customers. Individual performance measures

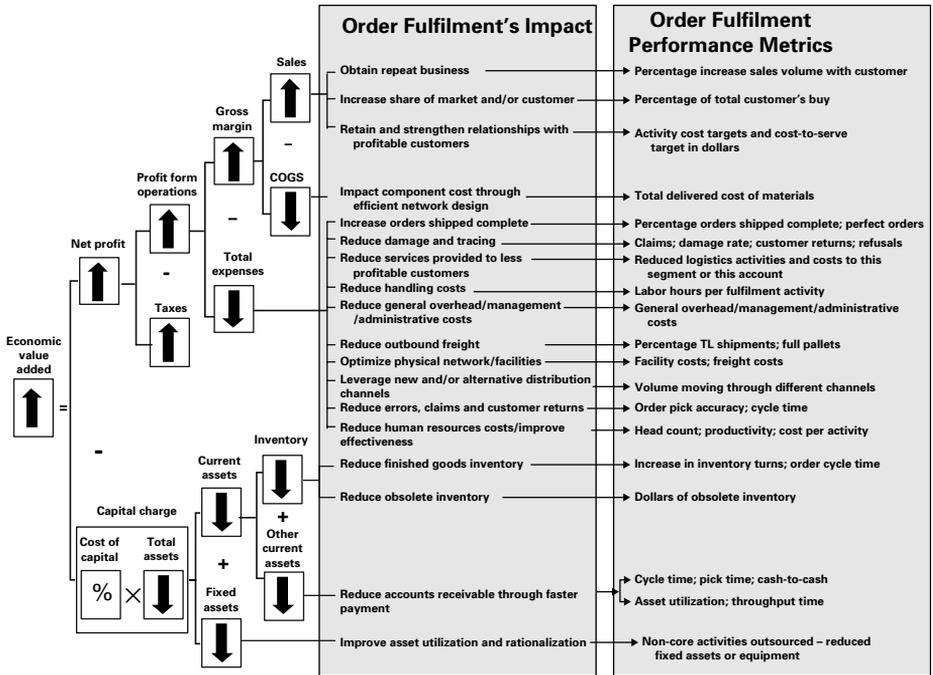


Figure 4.10: How order fulfilment affects economic value added
 Source: Lambert (2006, 85).

must be tied to the specific objectives required to improve profitability and shareholder value at each link in the supply chain. The relationship between improved operational performance and shareholder value can be accomplished by converting activities into cost, by identifying revenue or asset implications and then by inserting this information into an EVA[®] or profit analyses.

Compare across organizations and replicate

The final steps in the framework compare the resulting shareholder value and market capitalization (cap) across organizations (see figure 4.11) and replicate these steps at every link in the supply chain. Managers should assess whether the process changes and metrics employed have produced the targeted levels of profitability and shareholder value. They may need to refine the processes or make additional trade-offs to achieve the targets. In many instances, managers will need to look further upstream or downstream in the supply chain to achieve their objectives. They may find that second- or

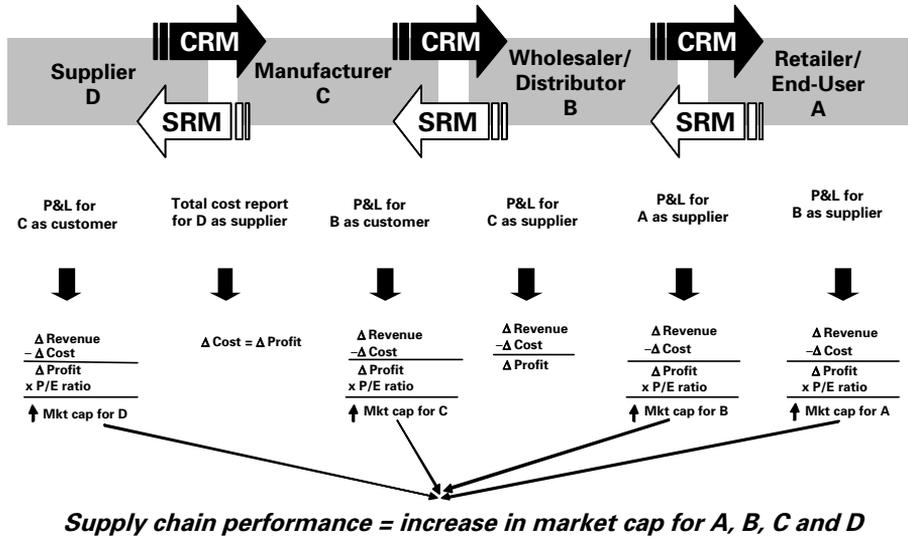


Figure 4.11: Profit and market capitalization increase measured across four tiers of a supply chain

Source: Lambert (2006, 117).

third-tier customers and suppliers provide additional opportunities to reduce cost, increase quality and accelerate product development. Management may decide to eliminate some intermediaries that do not add value or may insert others that can increase the profitability of certain segments – for example, using a distributor to service a large number of small accounts or to achieve distribution in a remote geographic region. It should be noted that there are other factors that can affect market cap, such as natural disasters or other external events, so it might be most practical to end with the total profit produced by the supply chain.

The supplier–customer profitability analysis should be applied at each link in the supply chain. By analysing the processes at each link and understanding the value the link creates, managers can align the supply chain processes towards the objectives that have the greatest effect on profitability and shareholder value for each company. This framework increases manager’s understanding of how their organization contributes to the overall competitiveness and value created by the supply chain. They can take proactive action within their organization as well as negotiate with other organizations to increase overall supply chain performance further. The most effective approach is to understand how value is created by each process at each link, take collaborative action to increase shareholder value and then replicate these steps across the entire supply chain.

This approach towards performance measurement of the supply chain may seem difficult, because most organizations have the ability to measure the profit impact of relationships only with tier 1 customers and suppliers. Several organizations in the quick-serve restaurant (QSR) industry, such as McDonald's and Wendy's, currently manage relationships and negotiate prices and service levels across multiple tiers of their supply chains. In these situations, it is possible for management to identify the total increase in profit for multiple tiers of the supply chain for certain products. This gives management the ability to drive performance towards the organization's objectives and maximize performance for the supply chain.

Conclusions

Executives are becoming aware of the emerging paradigm of inter-network competition, and that the successful integration and management of key supply chain management processes across members of the supply chain will determine the ultimate success of the individual enterprise. Managing the supply chain cannot be left to chance. For this reason, executives are striving to interpret and determine how to manage the company's supply chain network, and achieve the potential of SCM.

Research with member organizations of the Global Supply Chain Forum at the Ohio State University indicates that managing the supply chain involves three closely interrelated elements: (1) the supply chain network structure; (2) the supply chain business processes; and (3) the management components. The structure of activities/processes within and between organizations is vital for creating superior competitiveness and profitability. Successful SCM requires integrating business processes with key members of the supply chain. Much friction, thus wasting valuable resources, results when supply chains are not integrated, appropriately streamlined and managed. A prerequisite for successful SCM is the coordination of activities within the organization. One way to do this is to identify and manage the key business processes using cross-functional teams.

Most of the performance measures called "supply chain metrics" have an internal focus and do not capture how the organization drives value or profitability in the supply chain. These measures may actually prove dysfunctional, by optimizing an organization's performance at the expense of the other organizations in the supply chain – an approach that will eventually decrease the value of the entire supply chain. The use of customer

and supplier profitability reports (measured on revenue minus avoidable cost basis) reduces the probability of this occurring. The supplier–customer P&Ls will capture cost trade-offs as well as revenue implications, and the effect of any action taken by one organization will be reflected in both organizations' P&Ls.

The combined P&Ls provide the necessary foundation for improving performance in the supply chain. Although one organization may incur additional costs, the combined analysis will reflect whether the costs associated with a process improvement also generated increased profitability through a larger share of the customer's business or increased supply chain competitiveness. Replicating this approach and maximizing profitability at each link will align supply chain performance towards management's objectives and maximize performance for the whole.

References

- Anderson, D.L., Franke, E.B., and Donavon, J.F. (1997). The seven principles of supply chain management. *Supply Chain Management Review*, 1(1), 31–41.
- Beamon, B.M. (1989). Measuring supply chain performance. *International Journal of Operations and Production Management*, 19(3), 275–92.
- Cap Gemini Ernst and Young/Industry Week (2000). *High Performance Value Chains: A Report of the 2000 Value Chain Survey*. Paris: Cap Gemini Ernst and Young.
- Caplice, C., and Sheffi, Y. (1995). A review and evaluation of logistics performance measurement systems. *International Journal of Logistics Management*, 6(1), 61–74.
- Christopher, M.G. (1998). Relationships and alliances: embracing the era of network competition. In Gattorna, J. (ed.), *Strategic Supply Chain Management*, 272–84. Aldershot: Gower Publishing.
- (2005). *Logistics and Supply Chain Management: Creating Value-adding Networks*, 3rd edn. Harlow: Pearson Education.
- Cooper, M.C., Ellram, L.M., Gardner, J.T., and Hanks, A.M. (1997). Meshing multiple alliances. *Journal of Business Logistics*, 18(1), 67–89.
- Cooper, M.C., and Gardner, J.T. (1993). Good business relationships: more than just partnerships or strategic alliances. *International Journal of Physical Distribution and Logistics Management*, 23(6), 14–20.
- Cooper, M.C., Lambert, D.M., and Pagh, J.D. (1997). Supply chain management: more than a new name for logistics. *International Journal of Logistics Management*, 8(1), 1–14.
- Copacino, W.C. (1997). *Supply Chain Management: The Basics and Beyond*. Boca Raton, FL: St Lucie Press.
- Coyle, J.J., Bardi, E.J., and Novack, R.A. (2000). *Transportation*, 5th edn. Cincinnati: South-Western College Publishing.

- Croxton, K.L., García-Dastugue, S.J., Lambert, D.M., and Rogers, D.S. (2001). The supply chain management processes. *International Journal of Logistics Management*, 12(2), 13–36.
- Deloitte Consulting (1999). *Energizing the Supply Chain: Trends and Issues in Supply Chain Management*. New York: Deloitte Consulting.
- Drucker, P.F. (1998). Management's new paradigms. *Forbes Magazine*, 5 October, 152–77.
- ECR Performance Measures Operating Committee (1994). *Performance Measurement: Applying Value Chain Analysis to the Grocery Industry*, Joint Industry Project on Efficient Consumer Response. Watford: Institute of Grocery Distribution.
- Ellram, L.M., and Cooper, M.C. (1990). Supply chain management, partnership, and the shipper–third party relationship. *International Journal of Logistics Management*, 1(2), 1–10.
- Francella, K., and Doherty, K. (1998). *Supply Chain Solutions: Linking the Chains* (supplement prepared for Food Logistics). London: Andersen Consulting.
- García-Dastugue, S.J., and Lambert, D.M. (2007). Interorganisational time-based postponement in the supply chain. *Journal of Business Logistics*, 28(1), 57–81.
- Gilmour, P. (1999). A strategic audit framework to improve supply chain performance. *Journal of Business and Industrial Marketing*, 14(5–6), 355–63.
- Hammer, M., and Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*. New York: Harper Business.
- Hewitt, F. (1994). Supply chain redesign. *International Journal of Logistics Management*, 5(2), 1–9.
- Holmberg, S. (2000). A systems perspective on supply chain measurements. *International Journal of Physical Distribution and Logistics Management*, 30(10), 847–68.
- Kallio, J., Saarinen, T., Tinnila, M., and Vepsäläinen, A.P.J. (2000). Measuring delivery process performance. *International Journal of Logistics Management*, 11(1), 75–87.
- Keebler, J.S., Manrodt, K.B., Durtsche, D.A., and Ledyard, D.M. (1999). *Keeping Score*. Oakbrook, IL: Council of Logistics Management.
- Kuwaiti, M.E., and Kay, J.M. (2000). The role of performance in business process reengineering. *International Journal of Operations and Production Management*, 20(12), 1411–26.
- La Londe, B.J., and Pohlen, T.L. (1994). Issues in supply chain costing. *International Journal of Logistics Management*, 7(1), 1–12.
- Lambert, D.M. (2006). *Supply Chain Management: Processes, Partnerships, Performance*, 2nd edn. Sarasota, FL: Supply Chain Management Institute.
- Lambert, D.M., and Cooper, M.C. (2000). Issues in supply chain management. *Industrial Marketing Management*, 29(1), 65–83.
- Lambert, D.M., Emmelhainz, M.A., and Gardner, J.T. (1996). Developing and implementing supply chain partnership. *International Journal of Logistics Management*, 7(2), 1–17.
- Lambert, D.M., and Pohlen, T.L. (2001). Supply chain metrics. *International Journal of Logistics Management*, 12(1), 1–19.
- Langdon, D.S. (2001). Measure the whole. *iSource Business*, March, 33–6.
- Lapide, L. (2000). What about measuring supply chain performance? *Achieving Supply Chain Excellence through Technology*, 2, 287–97.

- Lee, H.L. (2000). Creating value through supply chain integration. *Supply Chain Management Review*, 4(4), 30–40.
- Lee, H.L., and Billington, C. (1992). Managing supply chain inventory: pitfalls and opportunities. *Sloan Management Review*, 33(3), 65–73.
- Levy, P., Bessant, J., Sang, B., and Lamming, R. (1995). Developing integration through total quality supply chain management. *Integrated Manufacturing Systems*, 6(3), 4–12.
- Macneil, I.R. (1980). *The New Social Contract and Inquiry into Modern Contractual Relations*. New Haven, CT: Yale University Press.
- Mentzer, J.T. (2001). *Supply Chain Management*. Thousand Oaks, CA: Sage Publishing.
- Mercer Management Consulting (undated). *Strategic Channel Management*, a Mercer Commentary. New York: Mercer Management Consulting.
- Neely, A.D., Gregory, M.J., and Platts, K.W. (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations and Production Management*, 15(4), 80–116.
- Olsen, R.F., and Ellram, L.M. (1997). A portfolio approach to supplier relationships. *Industrial Marketing Management*, 26, 101–13.
- Reese, A.K. (2001). Metrics mentality. *iSource Business*, June, 67–70.
- Rice, J.B., and Hoppe, R.M. (2001). Supply chain vs. supply chain: the hype and the reality. *Supply Chain Management Review*, 5(5), 46–54.
- Sherman, R.J. (1998). *Supply Chain Management for the Millennium*. Oak Brook, IL: Warehousing Education and Research Council.
- Simchi-Levi, D., Kaminsky, P., and Simchi-Levi, E. (2000). *Designing and Managing the Supply Chain*. Boston: Irwin McGraw-Hill.
- Stevens, G.C. (1989). Integration of the supply chain. *International Journal of Physical Distribution and Logistics Management*, 19(8), 3–8.
- Stock, J.R., and Lambert, D.M. (2001). *Strategic Logistics Management*, 4th edn. Burr Ridge, IL: McGraw-Hill Irwin.
- Supply-Chain Council (2000). *Supply-Chain Operations Reference Model: Overview of SCOR Version 3.1*. Pittsburgh: Supply-Chain Council.
- Towers, S. (1994). *Business Process Re-engineering: A Practical Handbook for Executives*. Cheltenham: Stanley Thorns.
- van Hoek, R.I. (1998). Measuring the unmeasurable: measuring and improving performance in the supply chain. *Supply Chain Management*, 3(4), 187–92.
- Walker, W.T. (1999). Use global performance measures to align the enterprise trading partners. *Achieving Supply Chain Excellence through Technology*, 1, www.ascet.com.
- Wilding, R. (2002). The 3Ts of highly effective supply chains. *Logistics Solutions*, December, 17–20.
- Williamson, O.E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

5

Finding performance: the new discipline in management

Marshall Meyer

Introduction

Performance is not an easy subject. There is a clear need to study and rethink what is meant by the performance of the firm and how to measure it. Performance has become the mantra of recent years. Many firms claim to be running for performance and seek to measure their performance, improve performance and compensate their people for performance. Yet, at the same time, there is widespread dissatisfaction with most performance measurement systems. Many firms, perhaps the majority, feel that they have not got it right. A 1995 article in *Chief Financial Officer* begins: “According to a recent survey, 80 percent of large American companies want to change their performance measurement systems...” The high level of dissatisfaction is sometimes attributed to the dearth of non-financial predictors of financial performance: “Yesterday’s accounting results say nothing about the factors that actually help grow market share and profits – things like customer service innovation, R&D effectiveness, the percent of first-time quality, and employee development” (Birchard, 1995). At the same time, controllers cite the burdens imposed by “newfangled performance measures” – read non-financial measures – as a key source of burnout, according to another article in *Chief Financial Officer* (Goff, 1995). Reports such as these, though anecdotal, suggest that executives are seeking measures that their controllers have so far been reluctant to deliver, leading to frustration on both sides.

Somewhat better evidence on the quality of measures comes from several surveys conducted by the cost management group of the Institute of Management Accountants (IMA), beginning in 1992 (Institute of Management Accountants, 1993, 1995, 1996). The coverage of the IMA surveys has increased over time. The 1992 survey covered 350 large US companies, but the 1996 survey included some 1,300 companies, ranging in size from approximately \$1 million to more than \$10 billion in sales; clearly, many

more small firms than large firms are represented in the IMA data. Most of IMA's informants are accountants, controllers, comptrollers and the like, although some are line officers. The bulk of the items in the IMA surveys asked companies what measures they now use and anticipate using in the future. In the 1992 survey 50 per cent of companies reported using non-financial measures, whereas 66 per cent of companies used non-financial measures in 1996. However, in 1992, 1993, 1995 and 1996 (there was no 1994 survey) 90 per cent or more of companies indicated that even greater use of non-financial measures would be desirable. Beginning in 1995, the IMA also asked companies whether they were measuring economic value added and whether they intended to implement EVA[®] at some future time. In 1995 18 per cent of companies were using EVA[®] and another 27 per cent were planning to use EVA[®]. By 1996 34 per cent of companies actually used EVA[®] and another 45 per cent anticipated using EVA[®].

Given their preferences for greater use of non-financial measures and their plans to implement metrics such as EVA[®], it is not surprising that measurement systems are changing more rapidly than even a few years ago. One of the key items in the IMA survey asked informants whether they had been involved in changing their company's performance measurement system. The proportion of managers responding affirmatively increased from 35 per cent in 1992 to 56 per cent in 1993, 64 per cent in 1995 and 63 per cent in 1996. (An item unique to the 1996 survey also asked managers to indicate whether they were undertaking "a major overhaul" of their current measures or replacing their entire performance measurement system; 60 per cent said they were.) What is surprising, however, is that many informants find that even the new measures do not adequately support the objectives of top management. The proportion of managers stating that their measurement system supports management objectives "less than adequately" or "poorly" rose unevenly in this period, moving from 35 per cent in 1992 to 43 per cent in 1993, 38 per cent in 1995 and 43 per cent in 1996. These changes are small and do not necessarily reflect a trend. However, they do suggest that, while people are changing their performance measures rapidly, they do not experience these changes as improvements.

In a world of perfect measurement, managers would be able to design optimal performance measurement systems. The measures of choice would meet the following requirements – note that I am not saying what the measures would be, only what the measures would look like:

- (1) There would be relatively few measures to keep track of, perhaps as few as three financial measures and three non-financial measures. This is a

matter of parsimony. If there are too many measures, cognitive limits will be exceeded and information will be lost.

- (2) The non-financial measures would predict subsequent financial performance – in other words, the non-financials would serve as leading performance indicators (and the financials as lagging indicators). Non-financials not demonstrated to be leading indicators would be sidelined, unless, of course, they were tracked as matters of compliance, ethics and security “must-dos” for firms.
- (3) These measures would pervade the organization – that is, the same measures would apply everywhere. Measures pervading the organization can be summed from the bottom to the top of the organization and decomposed downwards, the latter giving managers drill-down capability. Measures pervading the organization, moreover, permit performance to be compared across units.
- (4) The measurement system would be stable. Measures would evolve slowly, so as to maintain people’s awareness of long-term goals and consistency in their behaviour.
- (5) People would be compensated for performance on these measures – that is, for performance on both financial measures and the non-financial measures known to be leading indicators of financial results.

I wish I could recommend specific measures meeting these requirements. But I cannot. Such a measurement system, to the best of my knowledge, does not exist, and probably will not exist. Here is why.

First, firms are swamped with measures, and the problem of too many measures is, if anything, getting worse. It is commonplace for firms to have fifty to sixty top-level measures, both financial and non-financial. The longest list of top-level measures I have seen contains 117 measures – seventeen financial measures, seventeen customer measures, nineteen measures of internal process, thirty-five measures of renewal and development, and twenty-six human resources measures.¹ Many firms, I am sure, have even more top-level measures.

Second, our ability to create and disseminate measures has outpaced, at least for now, our ability to separate the few non-financial measures containing information about future financial performance from the many that do not. Some non-financial measures, such as customer satisfaction, properly measured, have been shown to predict financial performance,

¹ The full list of 117 measures is currently used by Skania, the Swedish financial services firm. See Edvinsson and Malone, 1997.

but the jury is still out on most measures (Anderson, Fornell and Mazvancheryl, 2004).

Third, it is very difficult to find non-financial measures that both predict financial performance and pervade the organization. It is somewhat easier to find financial measures that pervade the organization, but bear in mind that firms still struggle to drive measures of shareholder value from the top to the bottom of the organization.

Fourth, performance measures, non-financial measures especially, never stand still. With use they lose variance, sometimes rapidly, and hence the capacity to discriminate good from bad performance. This is the “use it and lose it” principle in performance measurement. The result is a continual shuffling of measures.

Fifth, compensating people for performance on multiple measures is extremely difficult. Paying people on a single measure creates enough dysfunctions, and paying them on many measures creates many more. The problem is combining multiple and often disparate measures into an overall evaluation of performance and hence compensation. If measures are combined formulaically, people will “game” the formula. If measures are combined subjectively, people will not understand the connection between measured performance and their compensation.

Here is the nub of my argument: the dissatisfaction that people experience with current measures and their yearning for better measures can be traced to an underlying but unrecognized cause. Superficially, the problem is measurement, and the solution is better measures. The measurement problem arises because the performance of the firm is not entirely measurable. Firm performance is, ultimately, future cash flows – “cash flows still to come” – discounted to present value (Fisher, 1988). Future cash flows cannot, by definition, be measured. What we can and do measure are past cash flows (financial performance), possible predictors of future cash flows (non-financial measures) are proxies for future cash flows (share prices). All these are imperfect measures – “second-best” measures, in my terminology. The measurement problem is one of finding the best of these second-best measures, sometimes through analysis, sometime through intuition. However, I don’t think measurement is the nub of problem. If measurement were the problem, then managers should be getting happier as their measures become more refined. Instead, they are getting more frustrated.

The more fundamental problem is that we are stuck with an archaic conception of the firm and where to look for performance. We think of firms mainly as black boxes. Investment flows into the box, activities take place

inside but out of sight, products are made and sold to customers as results of these activities, and an income statement, a balance sheet and a market valuation of the firm follow. Since financial results – the income statement, balance sheet and market valuation – are reported for the firm as a whole, or, internally, for large chunks of the firm called “business units”, we try to find measures describing the internal processes, products and customers of the firm or large chunks of it. These aggregate measures conceal important sources of variation within the firm. The things a firm does well are lumped together with the things it does poorly. Critical information about performance is obscured.

Activity-based performance measurement

To recover this information, to make the performance of the firm transparent, we must return to an elemental conception of the firm and what it does. Think of a firm as a bundle of activities, nothing more. These activities incur costs. These activities may also add value for the customer, although, equally, they may not. When activities add value for the customer, the customer supplies revenues to the firm. When activities do not add value, the customer holds on to his wallet. The elements of the firm, then, are activities, costs, the customer (who decides which activities add value and which do not) and revenues. The problem for the firm is finding those activities that add value for the customer and generate revenues in excess of costs, extending those activities, and reducing or eliminating activities that incur only costs. Finding performance is the issue – and, not coincidentally, the title of this contribution. Finding the right performance measures is a trivial task, although actually measuring performance, as will be seen, is not.

How did I come to the conclusion that an elemental conception of the firm is needed, that the problem is finding the activities that add value for the customer and generate revenues in excess of costs rather than finding better firm-level performance measures? I came to this conclusion mainly by struggling with anomalies, things that didn’t make sense. A few of these anomalies bear mentioning. One anomaly was definitional. Look in the *Oxford English Dictionary* under “performance”, as I did. You will find nothing about organizational performance. Performance is theatrical, mechanical or psychological. Performance, according to the *Oxford English Dictionary*, is what people or machines do. It is functioning, not an economic

result.² Other anomalies were empirical. Performance measures for firms are generally uncorrelated; this has been known for years.

This means that measurement is poor, raising the question of why firms pursuing performance would tolerate poor measurement, particularly when people's compensation depends on measured performance. Still other anomalies were in the response of the academic and business communities to my earlier work. Whenever I bashed existing performance measures I was received warmly; but I was also bashed whenever I suggested ways to improve current performance measures – with one exception. The exception occurred when I suggested that activity-based costing could be turned on its head to identify the drivers of revenues. This upending of ABC is called activity-based revenue, or ABR. Part of the appeal of ABR is its familiarity. It is, after all, based on ABC, an established technique. Part of the appeal of ABR, however, lies in the way it reduces the firm to activities, and the costs, customers and revenues associated with them. Reductionism is an established principle in science. Modern science reduces complex phenomena to simpler elements in order to understand and control them. The message I heard was that it is now time to apply reductionist principles to the management of firms, and the measurement of their performance in particular.

Modern performance measurement joins the dictionary definitions of “performance” and the prevailing definition of “economic performance”. The dictionary definition of “performance”, again, is the act of performing; what people, machines or, for that matter, firms do. The economic definition of “performance” is future revenues discounted to present value. *Modern performance measurement searches for what firms do that generates revenues in excess of costs.* However, having bridged the dictionary and economic definitions of “performance”, modern performance measurement has gone awry because it remains firm-centric. Firm-centric measurement treats the firm as a single entity and attempts to measure both the financial and non-financial performance at the level of the firm. It starts, in other words, with the firm and its financial results, asks how the functioning of the firm affects these results and then searches for the right measures of the functioning of the firm – that is, the right non-financial measures. This approach, I believe, is inherently flawed, because the right non-financial measures are hard to find and are always in dispute – recall the unhappiness that managers

² Readers are more likely to have access to the dictionary in Microsoft Bookshelf® than in hard copy. They will find roughly the same definitions of “performance”, and organizational performance is not mentioned.

experience with their measures. Activity-centric performance measurement, by contrast, decomposes the firm into the activities it performs and then identifies the costs incurred and revenues generated by each of these activities. Activity costs can be measured directly, and the revenues generated by activities can be measured indirectly provided that the activities performed for each customer and the revenues contributed by each customer are known. The principle difference between the firm-centric and activity-centric performance measurement, then, lies in the unit that is assumed to perform: the firm versus the activity.³

The implications of this difference for the quality of performance measures are profound. Firm-centric performance measurement measures everything. It creates a panoply of non-financial measures that may or may not contain information about the performance of the firm. There is no way to know without conducting elaborate statistical tests, which as often as not are inconclusive. Activity-centric performance, by contrast, requires a dictionary of the activities performed by a firm, measurement of the costs incurred and revenues generated by each of these activities, and maintenance of these measures. Creating an activity dictionary, measuring costs and revenues associated with activities and maintaining measures are daunting tasks, to be sure. Nonetheless, activity-centric performance measurement has the advantage of making the financial results of activities transparent. Activity-centric performance measurement thus avoids the problems created by a myriad of non-financial measures, the relevance of which to the bottom line is always in doubt.

My research on performance measures has identified some additional reasons firm-centric performance measurement is so frustrating to managers. One source of frustration is that firm-centric measures never stand still. They are always in turmoil. A key source of this turmoil lies in a phenomenon I call the “running down” of performance measures. Almost all measures lose variance, and hence the capacity to discriminate good from bad performance, as they are used. This triggers an ongoing search for new and different measures. Elsewhere, I illustrate “running down” by drawing on examples from major league baseball (specifically, batting averages), health care, nuclear power, commercial banking, mutual funds and the

³ Some will argue that there is a market-centric view of performance as well, because it is desirable to drive market-like measures into the firm. But I think not. The firm and the activity are *units on which performance is measured*. Financial markets value firms, and such valuations may or may not be useful indicators of the performance of the firm depending on their purpose.

J.D. Powers ratings of the quality of new cars (Meyer and Gupta, 1994). I also identify several causes of diminished variance in performance outcomes, among them positive learning (improvement), perverse learning or gaming (learning how to meet the measure without improving the performance that is sought), selection (replacing low performers with high performers) and suppression (withholding performance data when differences persist). While the causes of running down are different and often indistinguishable, the effects of running down are consistent, principally the ongoing pursuit of new and different measures. There is, in other words, a “use it and lose it” principle in performance measurement: the longer a measure is used and the more intensely it is applied, the less information it yields and the more urgent it becomes to find new measures.

Another source of frustration with firm-centric measurement is that it makes it very difficult to compensate people for measured performance. Many businesses have tried to pay their people using a combination of the financial and non-financial measures suggested by the “balanced scorecard”. The Western region of a firm I call “Global Financial Services” (GFS) tried to compensate its people on both financial and non-financial measures from 1993 to 1996. GFS’s initial approach to balanced compensation made use of a formula assigning explicit weights to various financial and non-financial measures. The intent was to place financial and non-financial performance on equal footings. There were unanticipated consequences, however. People learned how to game the formula and earn substantial bonuses without delivering bottom-line results – they were writing their own bonus cheques. In response, GFS’s management attempted to restore balance by adding contingencies that made bonuses payouts dependent on meeting certain financial and customer satisfaction hurdles. This complicated the bonus formula and made it very difficult to communicate.

GFS then switched to a compensation system in which financial and non-financial measures were weighted subjectively. Again, there were unanticipated consequences. In the absence of fixed weights, combining performance on different measures into an overall performance evaluation and a bonus payout was extremely time-consuming. People were, nonetheless, dissatisfied with the compensation system because they could not understand how they were being paid and whether they were being paid on the right measures. (It turns out that they were being paid on the wrong non-financial measures.) Moreover, despite the rhetoric of balance, little balance between financial and non-financial measures remained after several quarters of trying to weight financial and non-financial measures subjectively. Instead,

overall performance evaluations and compensation were increasingly determined by financial results – precisely the outcome GFS had sought to avoid (Ittner, Larcker and Meyer, 2003).

Activity-centric performance measurement starts from activity-based costing. A cardinal principle of ABC is this: if products or services are made to specifications known to add value for the customer, then activities, and hence costs, that can be removed without compromising these specifications are unnecessary and should be removed. This principle is responsible for many of the productivity improvements that have occurred in manufacturing. I then ask the following question: can performance be improved in complex service settings in which the specifications adding value for the customer are not known, or, more precisely, where the activities incurring costs cannot be easily separated from the specifications adding value? (Consider, for example, an airline journey in which the relative contributions to customer value of the cabin attendant's smile, the quality of the peanuts and an on-time arrival are not known.) Two studies I have conducted are germane to the problem. The first study compares a local competitor with Global Financial Services' retail operations in an Asian city. The local competitor succeeds by setting rigid specifications for customer service – customers' preferences are known, and the service specifications are surrogates for revenue drivers – and then removing unnecessary costs relentlessly. GFS cannot, however, imitate the strategy of standardizing service and removing costs because it is a global firm with customers that have different requirements. The second study reports the results of GFS's search for analytical methods to separate cost drivers from revenue drivers. The search originated in an ABC project in Latin America that nearly backfired because the cost-cutting recommendations made by the ABC team would have damaged the business if implemented. Rather than abandoning ABC, however, GFS transformed ABC into activity-based revenue, a tool capable of estimating the revenue consequences of customer transactions.

In the limited space allowed for this chapter, I cannot fully describe the implementation of ABR, nor can I illustrate the underlying concepts with graphics, but some of the basics can be sketched. First, transaction counters recorded in real time virtually all financial and non-financial customer transactions, the latter including balance enquiries, requests for documents and the like. The counters recorded virtually all the transactions taking place between each customer and the GFS organization. Overall, more than 400 types of transactions were tracked for more than 100,000 customers. Second, activity-based costing was carried out throughout the GFS organization that

was the subject of this study. Costs were identified at four levels: short-term variable, long-term variable, capacity and fixed costs. The first three categories of costs accounted for about 65 per cent of the organization's total expenditures. The unit cost of each type of transaction was then computed, based on the activities involved in the transaction initiated by the customer and the support transactions incident to it; note that there was no one-to-one correspondence of activities with transactions. Third, the revenues contributed by each customer were also recorded in real time. The three kinds of data available for each customer, then, were transaction frequencies by type, transaction costs and revenues. These data were reported monthly. In principle, it is a fairly simple matter to estimate both the short- and long-term revenues attributable to each transaction. Since transaction costs are known, the short- and long-term profitability of each transaction can be estimated as well. (This implementation of ABR is really transaction-centric rather than activity-centric, but keep in mind that all transactions are supersets of activities.)

The implications of ABR for the compensation of individual performance and the design of the firm are substantial. First, ABR drives individual accountability for results much deeper into the organization than firm-centric measurement. Some of the thorniest issues surrounding compensation disappear, although other issues remain. Second, as accountability is driven deeper into the organization, some of the classic dilemmas of organizational design recede in significance. Indeed, ABR renders individual accountability independent of organizational design. The result is that functional stovepipes that in the past have had to sacrifice accountability in order to achieve scale economies become more advantageous. The scale at which firms can operate effectively may be enhanced by ABR. ABR also blurs the distinction between the human resources and marketing functions, because the two share databases and methodologies: revenues at the activity (or transaction) level are estimated by modelling the revenue contributed by each customer as a function of the frequency of activities (or transactions) performed for that customer.

ABR does have substantial limitations, of course. It is most useful when many products are supplied to many customers and the product specifications adding value for customers are not understood – in other words, in complex service firms engaged in mass customization. ABR adds little when product specifications exist and are known to add value for customers – for example, in mass production (consider, for instance, DRAMs, for which the critical specifications are capacity, reliability and speed). Similarly,

ABR is not feasible when a wide range of products is supplied to a small group of customers, which would exhaust the degrees of freedom in revenue equations.

Considerable discipline is needed to implement activity-centric performance measurement and ABR in particular. The impetus for activity-centric performance measurement and ABR usually comes from experience with customers at both ends of the spectrum. Firms often find that their largest customers are among their most profitable *and* their least profitable customers; the largest customers can be highly profitable, because of the volumes involved, but they can also be highly unprofitable, because they negotiate the best prices and then consume inordinate amounts of the firm's resources (Kaplan and Cooper, 1987). Likewise, firms seeking to grow by adding small customers must understand their costs and the relationship of revenues to costs for these customers, or risk massive losses. The resistance to activity-centric performance measurement and ABR comes from accountants and financial controllers who find the approach unduly complex and imprecise, which, like any innovation, is in its formative stages. My experience has been that this resistance can be overcome by asking one question: "What is the alternative once all the low-hanging fruit has been picked, once conventional cost control methods have been exhausted?" So far, no one has come up with an alternative.

Conclusions

I began by echoing the complaints that people have about performance measures. These complaints are deeply felt. New and more refined firm-centric measures will not relieve these complaints any better in the future than they have in the past. A solution to the problem, I believe, lies in adopting an elemental view of the firm and in shifting performance measurement from the firm to the activity (or, in the case cited above, the transaction). By shifting attention from the firm to the activity, by shifting from firm-centric to activity-centric measurement and by implementing techniques such as ABR, performance becomes what the firm does and performance measures describe the costs and revenues resulting from this performance.

Measurement of this sort is not easy, but the quarrelling about what to measure vanishes. The critics will object to my looking inside rather than outside the firm to find performance. They will say, "You have failed to

consider that the performance of the firm is its return to shareholders.” My reply will be straightforward. The performance of a firm is what it does. If the firm performs well and if the firm adds value to customers in excess of its costs, and promises to continue to do so, then it will have performed well for its shareholders. Whether shareholders will profit from this performance is a somewhat different matter, because the wealth of shareholders also depends on the vicissitudes of the market.

References

- Anderson, E.W., Fornell, C., and Mazvancheryl, S.K. (2004). Customer satisfaction and shareholder value. *Journal of Marketing*, 68(4), 172–85.
- Birchard, B. (1995). Making it count. *Chief Financial Officer*, 11(10), 42.
- Edvinsson, L., and Malone, M.S. (1997). *Intellectual Capital: Realizing Your Company's True Value by Finding its Hidden Brainpower*. New York: HarperCollins.
- Fisher, F.M. (1988). Accounting data and the economic performance of the firm. *Journal of Accounting and Public Policy*, 7, 233–60.
- Goff, J. (1995). Controller burnout. *Chief Financial Officer*, 11(9), 60.
- Institute of Management Accountants (1993). *Cost Management Update* no. 32. Montvale, NJ: Institute of Management Accountants, Cost Management Group.
- (1995). *Cost Management Update* no. 49. Montvale, NJ: Institute of Management Accountants, Cost Management Group.
- (1996). *Cost Management Update* no. 64. Montvale, NJ: Institute of Management Accountants, Cost Management Group.
- Ittner, C.D., Larcker, D.F., and Meyer, M.W. (2003). Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. *Accounting Review*, 78(3), 725–58.
- Kaplan, R.S., and Cooper, R. (1997). *Cost and Effect: Using Integrated Cost Systems to Drive Profitability and Performance*. Boston: Harvard Business School Press.
- Meyer, M.W., and Gupta, V. (1994). The performance paradox. In Staw, B.M., and Cummings, L.L. (eds.), *Research in Organizational Behavior*, 309–69. Greenwich, CT: JAI Press.

6

A conceptual and operational delineation of performance

Michel Lebas and Ken Euske

Introduction

In this chapter the authors address the questions of what performance is and how to create it. The authors develop a series of nine propositions that, taken together, provide an answer to these questions.

After a brief overview of the reasons that led to these questions, and a review of the relevant literature that shows the diversity of meanings of “performance”, the authors develop step by step the process that leads to performance, showing it to be a social construct that results from the identification and the sharing of a causal model. That observation leads to the conclusion that performance is meaningful only within a decision-making context. The concept of performance is, therefore, specific to a given set of decision makers. Creating alignment between decision makers both inside and outside the firm is a prerequisite for performance to occur.

In the last sections of the contribution, the authors show the impact of responsibility assignment and of measurement on the operational definition of performance. All in all, the nine propositions form the basis on which performance can be defined, identified, measured and managed.

Performance

The word “performance” is widely used in all fields of management. In the management control area, terms such as “performance management” (Euske, Lebas and McNair, 1993), “measurement”, “evaluation” (e.g. Bruns, 1992) and “appraisal” are used. Despite the frequency of use of the word, its precise meaning is rarely explicitly defined by authors, even when the main focus of the article or book is performance (e.g. Baird, 1986; Richard, 1989). Often, performance is identified or equated with effectiveness and efficiency (e.g. Neely, Gregory and Platts, 1995; Corvellec, 1994). A publication of the

French Ministry of Industry (Ministère de l'Industrie, 1993) equated performance with lean production, competitiveness, cost reduction, value and job creation, growth and the long-term survival of enterprises. In short, “performance” is one of those “suitcase words” (Bourguignon, 1995) in which everyone places the concepts that suit them, letting the context take care of the definition.

Using context to clarify the meaning may help create a basis for understanding and discussion, but it may also engender ambiguous definitions. Ambiguity can be beneficial. Differing interpretations of the same reality may generate interaction that spawns new and creative outcomes. However, differing interpretations of the same reality may also generate interaction that is counterproductive and wastes scarce resources. The role of management and the systems they design is not to waste resources but, rather, to use the scarce resources to create value for the various stakeholders of the organization. If managers are to induce performance directly or through their systems and minimize counterproductive behaviour, they must know what performance is and what it implies.

A diversity of meanings of “performance”

A review of dictionaries (both French and English) shows a diversity of meanings for the term “performance”. It seems logical in the first place to list all these connotations, as their sum might provide a usable definition. Performance is:

- (1) measurable by either a number or an expression that allows communication (e.g. performance in management is a multi-person concept);
- (2) to accomplish something with a specific intention (e.g. create value);
- (3) the result of an action (the value created, however measured);
- (4) the ability to accomplish or the potential for creating a result (e.g. customer satisfaction, seen as a measure of the potential of the organization for future sales);
- (5) the comparison of a result with some benchmark or reference selected – or imposed – either internally or externally;
- (6) a surprising result compared to expectations;
- (7) acting out, in psychology;
- (8) a show, in the “performing arts”, that includes both the acting or actions and the result of the actions as well the observation of the performers by outsiders; and

(9) a judgement by comparison (the difficulty here is to define who the “judge” is, and to know on which criteria the judgement will be formed).

While Baird (1986) states that performance is action-oriented (i.e. it must be expressed by a verb), as opposed to a substantive or a noun that would refer to performance as an event, “performance” is referred to in most of the references as either an action (obtaining performance) or an event (a result), or both simultaneously. This list leads us to agree with Corvellec (1994, 1995) and Bourguignon (1995) in saying that performance refers simultaneously to the action, to the result of the action and to the success of the result compared to some benchmark. Viewing performance as a comparative judgement captures some of this complexity. If there is to be a judgement, a judge must be selected, and criteria for the judgement need to exist.

The criteria for the judgement are likely to focus on results, since the purpose of management is to create a continuous flow of value. Therefore, it becomes important to create a definition that will focus managers on the anticipation of performance. We take the position that performance is the sum of all the processes that will lead managers to taking appropriate actions in the present that will create a *performing* organization in the future (i.e. one that is effective and efficient). In other words, we define “performance” as doing today what will lead to an outcome of *measured* value tomorrow.

To create something in the future a causal model is necessary, so that the process through which performance (future results) is to be created can be identified and managed. Past performance (past results) alone is not necessarily a good predictor of future performance. There are many illustrations of the lack of predictability of results, and very few examples of predictable results.

Performance and the causal model

A causal model that links actions now to results in the future can take a variety of forms. Figure 6.1 illustrates an example of a generic three-stage causal model consisting of:

- outcomes (often reduced to output and results);
- processes; and
- foundations.

Each firm or organization will need to define uniquely the concepts that apply to its own situation. The very process of defining the three components

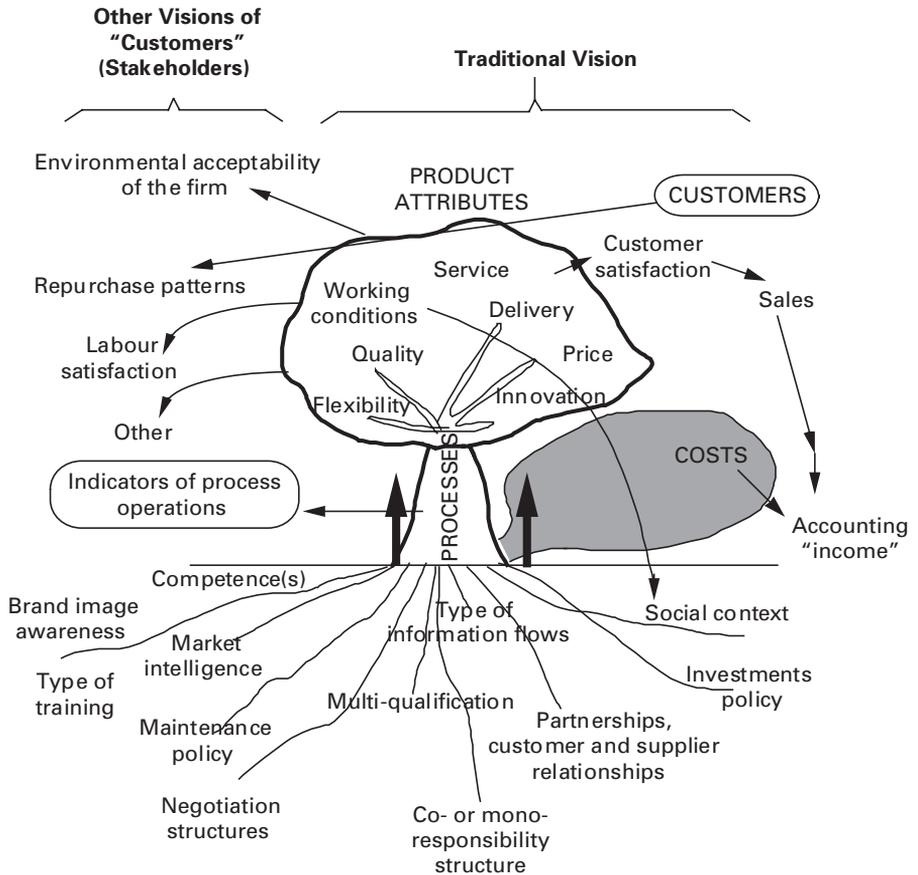


Figure 6.1: The performance tree

Source: Adapted from Lebas (1995).

of the model is, in our view, an essential step in creating performance. Once the model is defined, each organization must select the appropriate indicators¹ to describe it and monitor its status.

This model is portrayed as a tree to illustrate how an organization goes through the process of creating performance. The analogy to a tree helps to capture process complexity and the characteristics of growth and change.

¹ We deliberately prefer the word “indicator” to the more traditional one of “measure”. A measure often implies precision; it is usually well defined, and in similar circumstances its numerical value should be the same. An indicator may be less precise, but meaningful; indicators tend to allow for more timely and sensitive signals.

In our illustration outcome, results, or outputs, are divided into two broad categories: traditional conceptualizations and other conceptualizations. Accounting income, shown on the right-hand side of the figure, is an example of a traditional conceptualization of a result that might be held by an owner-manager or a stockholder. However, other results are valued by groups of stakeholders, such as the environmental acceptability of the organization and its contribution to the social welfare, labor and social climate. Social climate is particularly important, because it captures the continued acceptability of the organization to the political, regulatory and administrative powers that, either implicitly or explicitly, grant the organization its licence to operate (Fligstein, 1990; RSA, 1995).

These outputs are consequence of the product attributes that constitute the fruit of the tree. These attributes are the elements of the product that the customer values. They include, of course, the traditional quartet: price, availability, service and quality. They can also include other elements, such as working conditions (e.g. buying union-made products or not buying products made by child labour), innovation and flexibility. The attributes are the basis for customer satisfaction, but also for stakeholder satisfaction in general. The attributes are the result of business processes, which constitute the trunk of the performance tree. They have to be monitored so that they deliver what the stakeholders want within the constraints of the strategic intent of the organization.

Costs that loom quite large as descriptors of financial performance, either directly (cost minimization) or indirectly (earnings maximization), do not play a large role in our causal model; costs are the mere “shadow” of the processes and of the attributes created.² Costs are important but they are second-order variables in the understanding of the generation of results.

Continuing the tree analogy, the quality of the processes would be the richness of the sap and its effective movement through the trunk and branches. Further, the quality of processes rests in part on the nutrients in the soil. They are such elements as competence, awareness of brand value, maintenance policy, existing structures of negotiation, partnerships with both customers and suppliers, and the organizational responsibility structure – concepts that are not normally captured in accounting and control

² Johnson (1990) has used Plato’s cave analogy to explain the concept. In the cave analogy, above ground are objects and forms such as customer satisfaction and activities. Below ground are artificial likenesses and shadows such as accounting information and costing models.

systems. Why are these elements rarely measured by these systems? They are in the “soil”, and do not normally catch the light that is needed to create the accounting “shadow”. If they do catch the light (i.e. are “seen” by the accounting system) it is a sign that things are not going very well, because it signifies that the tree has been uprooted.

Modelling the performance creation process as a tree offers an opportunity to visualize that outcome results, or outputs, often do not occur in the same time frame as that of actions: the work in the soil, the choice of the type of tree and caring for the tree are all actions that have to be implemented long before any fruit can be seen, let alone harvested. Just as a tree takes several years to bear fruit, the consequences of the interaction with the environment (e.g. the bad health of the workers or of the surrounding community due to hazardous chemicals) take time to materialize. Accounting data does not capture such lags. This illustration is consistent with the fact that the results of an organization are multifaceted and must be described over a long period of time.

Performance is a complex concept. The complexity increases both the difficulty of defining the concept and the likelihood that indicators of performance will at times be contradictory. The contradictions can be managed if one has a good understanding of the process that generates the various types of results; hence the importance of the causal model as a means to understand the organization and its interaction with its environment. However, once a model is adopted, performance, because we define it as the process as well as the future outcome, cannot be separated from the model. The model both defines and legitimates the performance (Fligstein, 1990). Performance is a social construct. The model creates the reality as to what performance is.³

The causal model is elaborated by trial and error, critically using past data.

Proposition 6.1

Performance can be expressed only as a set of parameters or indicators that are complementary, and sometimes contradictory, and that describe the process through which the various types of outcome and results are achieved (Lebas, 1995; Kaplan and Norton, 1992).

³ The view taken in this chapter can be illustrated by a story told about the definition of a “penalty” in football, in which the first referee says: “I blow my whistle when *there is* a penalty,” the second says: “I blow my whistle *when I see* a penalty,” while the third one declares: “There is no penalty *until I blow my whistle*.”

Proposition 6.2

Understanding performance relies on the identification of a causal model that describes how actions today can influence results in the future. Performance is not a one-time event. Performance is dynamic. A performance measure is an instance in the continuous performance creation process. A performance measure is a leading indicator of performance only if the organization has acquired the knowledge and the mastery of its causal relationships and can reproduce this outcome or result in the future. We suggest that the term “performance” be reserved for the sum of all processes that lead to a potential or future sequence of outcomes and results.

Performance and decision making

Even with a common causal model, the description of performance – whether simple, complex, cardinal, ordinal or literal – has no intrinsic value. The description becomes valuable if one or several individuals use it for decision-making purposes. If the description of performance has no possibility of impacting on the decision of the user, it has no value. However, each user can interpret the performance data as he or she pleases according to different time frames, objectives, intent, risk avoidance attitudes or perspectives (inside or outside the organization). This diversity of interpretation increases the complexity of providing a definition of performance.

A description of performance that would be correct from a fiduciary perspective (e.g. the balance sheet) is not likely to satisfy a user who views the organization as an operating entity. Such a description would likely be even less meaningful to a stakeholder preoccupied by the impact of the organization on the social welfare in a community. Even though the decision context may be the same, users with differing time horizons or differing objectives are not likely to seek the same description of organizational performance.

Given a common causal model, the perceived contradiction between the various views of performance may be related to a world-view based on the concept of “or”, implying a concept of exclusion (i.e. some may say the organization *either* is profitable *or* maintains employment). However, a view based on the concept of “and” implies a concept of inclusion (that is, the organization can be profitable *and* maintain employment). From an internal, operational point of view, the difference is between dissolving a

conflict and living with a conflict (i.e. accepting the coexistence of multiple dimensions of the same concept). The manager is no longer faced with a dilemma. Rather, the manager has the challenge of taking a proactive position regarding the complexity of performance. One consequence of this view is that, today, some organizations see profit as a constraint, not a goal per se.

Proposition 6.3

Performance is defined by the user of the descriptive signals of performance. Performance, because it is a social construct, is a concept with no objective description. Each person defines it her or his own way.

Performance defined from inside or from outside the organization

Someone inside or outside the organization can develop a causal model to define performance. An internally defined model of causal relationships is likely to focus on the construction of the result through actions. An externally defined model is more likely to focus on anticipating the possible actions the internal actors might select, and estimate the probability of certain future results to be used in some other decision-making process.

The descriptors used in the two cases will not be the same. An outsider will look at general indicators based on some preconceived, and possibly statistically defined, relations, such as the ones found in financial analysis. The actual workings of the organization will remain a black box. It will not be surprising, therefore, that performance as seen by outsiders will lead to much debate, as each analyst will necessarily introduce his or her own bias in the inferences drawn from externally available signals about the activity of the organization.

An insider, on the contrary, will model action variables. While it is normal to have a diversity of views about performance as seen from the outside, the concept of performance as defined from the inside of the organization is more likely to have a unique, although multifaceted, definition, shared by all actors involved in its creation. If the members of an organization do not share the same view of performance, actions cannot be coordinated and resources may be wasted.

Proposition 6.4

Performance does not have the same meaning if the evaluator is inside or outside the organization. The operations of the organization remain a black box for the outsider, while the insider operationalizes performance in cooperation with other internal actors.

Performance and responsibility

For reasons of effectiveness and because of limits on individual competence, each causal model is generally broken down into sub-models. The sub-models are either additive (a Taylor-based view, still prevalent in many organizations) or overlapping (viewing the enterprise as a network of cross-functional processes and management of the “white space”⁴ on the organization chart). Each sub-model defines a domain of responsibility (e.g. the manager, the responsibility centre,⁵ teams, the management systems, the product or service) and, conversely, each definition of responsibility implies a causal model.

For each of these domains of responsibility there will be different descriptors of performance, and different uses for the signals describing it. Not all these descriptors will necessarily be consistent with one another. Therefore, it is crucial to admit that performance does not have a unique operational definition in an organization. However, the dialogue that will take place to define the richness and the complexity of the concept will be a foundation for the management of performance (i.e. for the proactive construction of performance).

Proposition 6.5

Performance is always connected or attached to a domain of responsibility. The different views of performance associated with the domains provide the basis for an understanding of the complexity and management of performance in the organization.

⁴ The concept of “white space” is developed by Rumler and Brache (1990).

⁵ It is generally understood that it is important to separate the performance of an organizational subunit (generally a responsibility centre) from that of the individual(s) in charge of the subunit. A manager may very well have good performance in an organizational subunit that does not perform well. For example, a manager may do a great job of closing down a loss-making branch or subsidiary.

Performance and measurement

As Lord Kelvin once said, “If you cannot measure it, it does not exist.” As we have said, performance is multifaceted and encompasses elements describing both the results and the processes creating the results. However, the descriptors, the qualitative and quantitative measures, are mere surrogates of performance. They should not be mistaken for performance itself (Euske, 1983). Accounting definitions and measures of performance are but synthetic representations of decisions that were made previously by managers and that can be visualized as parts of the “performance tree” of figure 6.1. It is important for management accounting to identify, measure and transmit data about these intermediate results, even though they may not be expressed in “accounting language”. The management accounting process is a mechanism to provide legitimacy to what may be estimations or forecasts. These estimates and forecasts may be better descriptors of the process than accounting data.

Accounting data or quasi-accounting data are provided to the manager for his or her information. They need not be used if the causal model used does not require them. However, the causal model may not be permanently relevant, and sometimes intuitions derived from the data will lead to an update of the causal model to reflect the rapidly evolving markets and technologies. Therefore, it may be important to add to the complexity of measures or indicators, by recognizing that there will be two types of signals: those assuming that the model is still valid (efficiency and effectiveness, for example) and those allowing a verification of the continued relevance of the model.

Proposition 6.6

Performance exists only if outcome and results can be described or measured so that they can be communicated for someone to decide to do something within the shared model of causal relationships.

Proposition 6.7

The relevance of the causal model needs to be validated continuously, both within and without the organization.

Proposition 6.8

Performance indicators or measures should not be confused with what they only partially describe.

“Performance” is only a relative term

Performance corresponds to a potential for value creation. That value is to be created over a period of time. Any causal model must, therefore, specify a time frame, in addition to the decision parameters and a context. The context is comparative.

Performance measures and the underlying performance must be qualified as good or bad. No signal of performance is intrinsically either. There must always be a comparison to qualify the performance. If, for example, we consider that late deliveries are a parameter descriptive of an organization’s performance, we cannot decide whether 10 per cent of orders being five days late is good or bad. We cannot even decide whether a reduction of late orders from 10 per cent to 5 per cent is really an improvement or not. In order to interpret these data, one has to know (or surmise) what competitors or other users of similar processes do. Performance cannot be taken out of its comparative context, as shown in figure 6.2.

In this figure, the solid line indicates that, over time, there is an absolute improvement in the service rate, defined as the percentage of late deliveries.

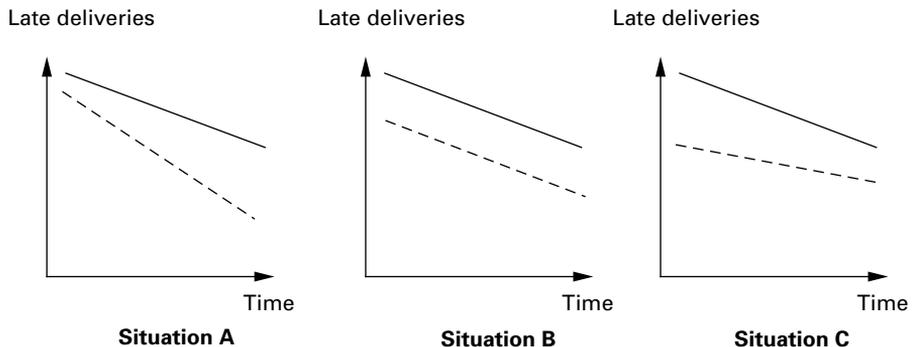


Figure 6.2: Performance is only relative

From a continuous improvement perspective, the change is clearly an improvement. From another perspective, however, the change in performance may not be a relative improvement. For instance, in the three situations described in figure 6.2, another organization using a similar or comparable process enjoyed a reduction of late deliveries shown by the dotted line. In situation A, the organization indicated by the solid line is losing ground relative to the other organization. In situation B, the relationship of the two organizations remains constant. Only in situation C is the gap between the two organization narrowing. The concept of performance only has meaning as part of a comparison.

Proposition 6.9

Performance is a relative concept, requiring judgement and interpretation. Performance is effecting a superior process or result relative to the referent. Choice of the referent is a significant decision with long-term consequences. The relatively superior position could be short- or long-term and over few or many indicators. Contradictions among the temporal measures and the other indicators are inevitable. Performance will again be in an interpretative context, in which managers or users of information will decide on the key parameters of performance.

Performance can be managed

As mentioned above, measuring parameters descriptive of performance makes sense only if the data are to be used in making decisions. The decisions can relate to both strategic orientations and steering the organization in the implementation of the strategic intent. The decisions contribute jointly to the creation (i.e. the management) of performance.

In order to achieve performance, the causal model has to be defined in terms of leading indicators. Lagging indicators provide only history; leading indicators allow for the creation of the conditions for fostering performance. In order to maintain the validity of the leading indicators, the model must be continuously validated for its relevance. This validation must also be as “leading” as possible (i.e. incorporate the most current information pertaining to the causal model and its parameters).

Performance management requires that procedures be put in place that allow the evolution of the organization and of its management system in

line with the evolution of its environment. Therefore, in order to manage performance one must:

- describe the value creation process in its context and time (propositions 6.1 and 6.2);
- share this model with all relevant actors (proposition 6.4);
- partition and allocate decision rights on the basis of this model (proposition 6.5);
- identify and select the descriptive indicators both for results and for steps to creating the results (propositions 6.3 and 6.4);
- document these indicators through an appropriate information system (proposition 6.6);
- choose the reference for benchmarking and external validation (propositions 6.7, 6.8 and 6.9);
- evaluate the signals and messages coming from each indicator (proposition 6.8); and
- identify, evaluate and implement all actions likely to improve the likelihood that the result will be coherent with the strategic intent (propositions 6.7, 6.8 and 6.9).

Conclusions

Performance is not just something one observes and measures; it is the result of a deliberate construction. Performance is a relative concept, defined in terms of some referent employing a complex set of time-based and causality-based indicators bearing on future realizations. Performance is about the capability of generating future results. The capability of generating future results can be described through a causal model. Each part of the model can, in turn, be subjected to an analysis.

Performance is meaningful only when used by a decision maker. It is specific to the individual's needs and interpretation. A domain of responsibility defines the parameters of performance that are relevant and, conversely, performance defines a domain of responsibility.

Finally, the specific meaning that performance takes in an organization should be the result of extensive discussions between the various managers or decision makers of the organization. The goal of the discussions is to identify a coherent set of causal relationships and select a common set of indicators so that coordination of all the actors takes place and generates

value such that, in the end, stakeholders define performance from their own point of view.

This definition of the performance creation process highlights the importance of creating alignment as a basic condition for an efficient use of resources and an effective trend towards the fulfilment of strategic intent.

Performance management is the process of creating alignment. Some of the best-known processes leading to such alignment are dialogue-based and de-emphasize local optimization, focusing on the development of integrated business processes

Figure 6.1 shows the conceptual three-step approach and highlights the fact that, unless foundations (positions, views and beliefs) are well understood and managed, outcome and results can hardly be modified.

References

- Baird, L. (1986). *Managing Performance*. New York: John Wiley.
- Bourguignon, C. (1995). Peut-on définir la performance? *Revue Française de Comptabilité*, 269, 60–5.
- Bruns, W.J., Jr. (1992). *Performance Measurement, Evaluation and Incentives*. Boston: Harvard Business School Press.
- Corvellec, H. (1994). Performance: from one language into another or the mutations of a notion. Paper presented at Seventeenth European Accounting Association Annual Congress, Venice, 6–8 April.
- (1995). *Stories of Achievements: Narrative Features of Organizational Performance*. Lund, Sweden: Lund University Press.
- Euske, K.J. (1983). *Management Control: Planning, Control, Measurement, and Evaluation*. Reading, MA: Addison-Wesley.
- Euske, K.J., Lebas, M.J., and McNair, C.J. (1993). *Best Practices in World Class Organizations*. Burleson, TX: Consortium for Advanced Manufacturing - International.
- Fligstein, N. (1990). *The Transformation of Corporate Control*. Cambridge, MA: Harvard University Press.
- Johnson, H.T. (1990). Beyond product costing: a challenge to cost management's conventional wisdom. *Journal of Cost Management*, 3(3), 15–21.
- Kaplan, R.S., and Norton, D.P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.
- Lebas, M.J. (1995). Oui, il faut définir la performance. *Revue Française de Comptabilité*, 269, 66–71.
- Ministère de l'Industrie (1993). *Guide de la performance industrielle*. Paris: Ministère de l'Industrie, Délégation à l'Information et à la Communication.

- Neely, A.D., Gregory, M.J., and Platts, K.W. (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations and Production Management*, 15(4), 80–116.
- Richard, J. (1989). *Audit des performances*. Paris: La Villeguérin.
- RSA (1995). *Tomorrow's Company: The Role of Business in a Changing World*. London: Royal Society for the Encouragement of Arts, Manufacture and Commerce.
- Rumler, G.A., and Brache, A.P. (1990). *Improving Performance: How to Manage the White Space on the Organization Chart*. San Francisco: Jossey-Bass.

PART II

Performance measurement – frameworks and methodologies

The second part of the book explores some of the frameworks and methodologies associated with performance measurement. While there is considerable interest in the balanced scorecard, there are, of course, numerous other measurement frameworks and methodologies, each with their own strengths and weaknesses.

The first contribution in this part, from Andy Neely, Mike Kennerley and Chris Adams, reviews some of these other measurement frameworks and then proposes an alternative framework: the performance prism. Neely, Kennerley and Adams argue that the strength of the performance prism lies in the fact that it unifies existing measurement frameworks and builds upon their individual strengths. The balanced scorecard, for example, is strong in that it argues for a balanced set of measures, but weak in that it omits some extremely important stakeholder perspectives – i.e. those of employees and suppliers. Similarly, activity-based costing is strong in that it explicitly recognizes the importance of activities and processes, but weak in that it does not link these processes back to strategies or stakeholders. The performance prism addresses these, and other issues, by providing an integrated framework with which to view organizational performance.

The second contribution explores the concepts of “beyond budgeting” and the adaptive organization. In recent years there has been significant interest in beyond budgeting, a notion promulgated by Jeremy Hope and Robin Fraser. In this contribution Jeremy Hope explores how the beyond budgeting framework might be applied to the adaptive organization, arguing that managers need to understand how complex systems make everything unpredictable and therefore have to: (a) manage through continuous planning cycles; (b) devolve planning to front-line teams; (c) make rolling forecasts the primary management tool; (d) report key metrics daily and

weekly; (e) manage resources dynamically; (f) coordinate cross-company actions dynamically; and (g) focus accountability on the relative performance of teams.

The third and fourth contributions – those from Hanne Nørreklit, Lennart Nørreklit and Falconer Mitchell and from Stephen Pike and Goran Roos – ask questions about the validity of measurement frameworks, from different perspectives. Nørreklit, Nørreklit and Mitchell explore the philosophical underpinnings of performance measurement and measurement frameworks, while Pike and Roos challenge measurement frameworks using the mathematically derived measurement theory.

Nørreklit, Nørreklit and Mitchell's chapter starts by observing that truth is fundamental to accounting, but then opens up the question of what "truth" means in the context of accounting. They point out that many accountants assume they are reflecting truth (in doing so they adopt a realist's position, assuming that the world is out there and can be objectively observed), yet in practice many accounting approaches are inherently subjective (e.g. the assessment of goodwill, the application of depreciation). Nørreklit, Nørreklit and Mitchell introduce pragmatic constructivism as an alternative way of conceptualizing the processes of accounting and explore the implications of this for designing performance measurement systems.

Pike and Roos adopt a different starting point for their contribution by drawing on measurement theory, a branch of applied mathematics. They make the important distinction between "value" measurement and "performance" measurement. They argue that organizations often want to use their measurement systems to make value judgements – how much extra is an X per cent increase in customer satisfaction worth? – but that such value judgements can be made only by using measurement systems that conform with measurement theory. In explaining the axioms underpinning measurement theory, Pike and Roos call into question the veracity of many of the measurement frameworks used today.

Performance measurement frameworks: a review

Andy Neely, Mike Kennerley and Chris Adams

Introduction

The shortcomings and dysfunctional consequences of performance measurement systems have been discussed in the academic literature for at least fifty years (Ridgway, 1956), but recently there has been a flurry of activity. Throughout the 1980s vocal and influential authors criticized the measurement systems used by many firms (Johnson and Kaplan, 1987; Hayes and Abernathy, 1980). By the 1990s the noise made by these voices had grown to a crescendo (Eccles, 1991; Neely, Gregory and Platts, 1995), and increasing numbers of firms appeared to be “re-engineering” their measurement systems, with data suggesting that, between 1995 and 2000, 30 to 60 per cent of companies transformed their performance measurement systems (Frigo and Krumwiede, 1999). By 2001 the balanced scorecard had been adopted by 44 per cent of organizations worldwide (57 per cent in the United Kingdom, 46 per cent in the United States and 26 per cent in Germany and Austria). More recent data suggest that 85 per cent of organizations should have had performance measurement system initiatives under way by the end of 2004 (Rigby, 2001; Silk, 1998; Williams, 2001; Speckbacher, Bischof and Pfeiffer, 2003; Marr *et al.*, 2004). However, cautionary evidence was reported by three Austrian academics to the effect that 8 per cent of 174 companies from German-speaking countries decided not to implement a performance measurement system (and a balanced scorecard in particular) because they could see no advantages or “positive impact” for themselves, especially given the implementation effort required (Speckbacher, Bischof and Pfeiffer, 2003).

In response to “calls from practice” for new and better ways of measuring organizational performance, the academic and consultancy communities

have developed a plethora of performance measurement frameworks and methodologies. The objective of this chapter is to review these performance measurement frameworks and methodologies and identify the key characteristics that they exhibit. Building on these, we seek to develop a new performance measurement framework – the performance prism – that incorporates the best of the existing frameworks and methodologies (Neely, Adams and Kennerley, 2002). Since its introduction the performance prism has been applied in various settings, and the chapter ends by illustrating some of the lessons learned in the years since it was first introduced.

Existing performance measurement frameworks

The history of performance measurement – particularly accounting performance measurement – can be traced back through the centuries. Some of the earliest references are found in the writings of Venetian monks, who invented the double-entry bookkeeping system in the thirteenth century. Other references – to the problems of performance measurement – can be found in the writings of W. Hamilton Church, who pointed out in 1908 one of the shortcomings of the overhead allocation processes:

Shop charges (overhead) frequently amount to 100 percent, 125 percent, and even much more of the direct wages. It is therefore actually more important that they should be correct than that the actual wage costs should be correct. (quoted in Kaplan, 1984, 395)

In 1951 the General Electric Company established a “Measurement Project” that “was intended to develop performance metrics that could be applied on a decentralised basis” (Meyer and Gupta, 1994, 348). It is rumoured that the same project was still running twelve years later. Johnson (1992) argues that “management by remote control” – i.e. managing by the financials – became popular only after the 1950s, before which time senior managers used the financial figures for planning rather than for control.

For many years frameworks have been used by organizations to define the measures that they should use to assess their performance. From early in the twentieth century DuPont used a pyramid of financial ratios, which linked a wide range of financial ratios to return on investment (Chandler, 1977). The pyramid of financial ratios has an explicit hierarchical structure that links measures at different levels of disaggregation (see figure 1.1). One of its great

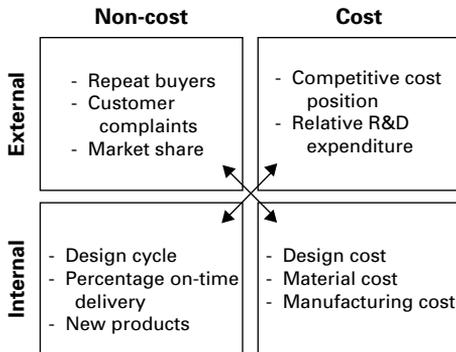


Figure 7.1: The performance measurement matrix

Source: Keegan, Eiler and Jones (1989).

strengths is that it makes explicit the “levers” that management can pull as they seek to influence performance. Nonetheless, the pyramid of financial ratios – along with many other purely financial measurement systems – has been criticized for overemphasizing historical financial performance and encouraging short-termism (Banks and Wheelright, 1979; Bruns, 1998).

As these criticisms gathered pace, managers began the search for new and better methods of measuring organizational performance. Many of these were designed to enable the adoption of non-financial performance measures to supplement the existing financial measures, with the aim of better capturing the breadth of the organization’s objectives. This search for so-called “balanced” measurement systems led to the development of a plethora of measurement frameworks designed to help organizations implement balanced sets of measures.

The early frameworks

One of the earliest frameworks that gained reasonably widespread recognition was the performance measurement matrix (PMM) of Keegan, Eiler and Jones (1989). The performance measurement matrix categorizes measures as being “cost” or “non-cost”, and “internal” or “external”. This simple design is supposed to reflect the need for more balanced measurement systems (see figure 7.1), and its inherent flexibility means that it can accommodate any measure of performance (Neely, Gregory and Platts, 1995).

The SMART (strategic measurement and reporting technique) pyramid developed by Wang Laboratories (Lynch and Cross, 1991) also supports the

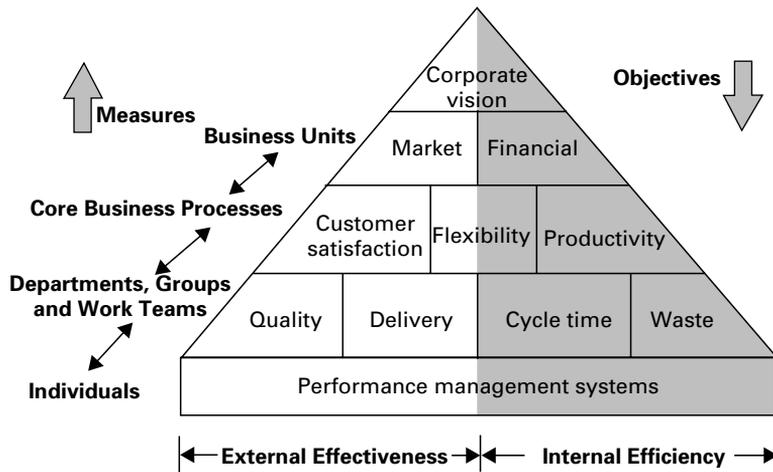


Figure 7.2: The SMART pyramid

Source: Lynch and Cross (1991).

need to include internally and externally focused measures of performance. It adds the notion of cascading measures down the organization, so that measures at departmental and work centre level reflect the corporate vision as well as internal and external business unit objectives (see figure 7.2).

Following their study of performance measurement in service industries, Fitzgerald *et al.* (1991) proposed a framework classifying measures into two basic types: those that relate to results (competitiveness, financial performance) and those that focus on the determinants of those results (quality, flexibility, resource utilisation and innovation). A particular strength of the results–determinants framework is that it reflects the concept of causality, emphasizing that the results obtained today are a function of past business performance in relation to specific determinants. In essence, the results–determinants framework encapsulates the theme raised in later debates about designing and deploying performance measurement systems – namely the need to identify drivers of performance in order to achieve the desired performance outcomes (see figure 7.3).

Brown (1996) further developed the concept of linking measures through cause and effect relationships. In his macro process model of the organization, he shows clear links between five stages in a business process and the measures of their performance. These stages are defined as inputs, processing system, outputs, outcomes and goal respectively. Brown's model assumes a linear set of relationships between inputs, processes, outputs, outcomes and goals, with each previous factor determining the next. While this is, clearly,

Results	Competitiveness
	Financial performance
Determinants	Quality
	Flexibility
	Resource utilization
	Innovation

Figure 7.3: The results–determinants framework

Source: Fitzgerald *et al.* (1991).

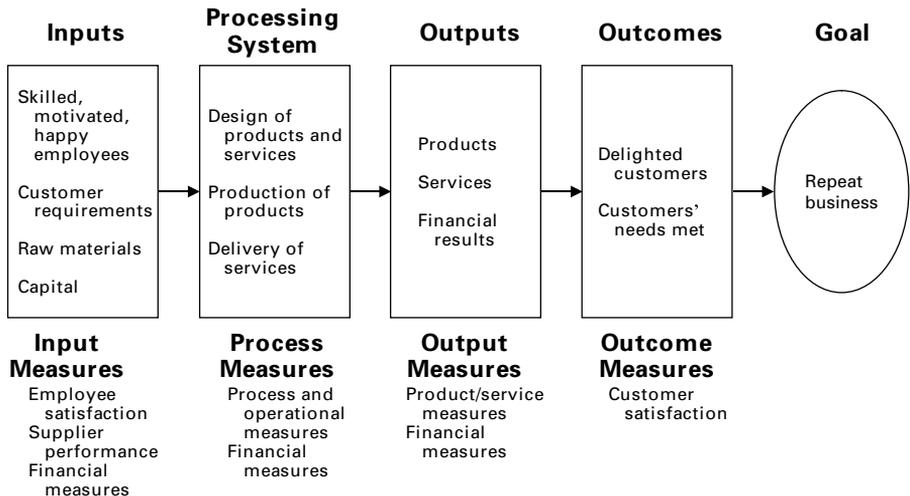


Figure 7.4: The input–process–output–outcome framework

Source: Brown (1996).

an oversimplification of reality, the model is a useful way of distinguishing between different categories of measures (see figure 7.4). The distinction between output and outcome measures has proved particularly popular in the public sector.

The most popular of the performance measurement frameworks has been the balanced scorecard, proposed by Kaplan and Norton (1992, 1996a). The balanced scorecard identifies and integrates four different perspectives in terms of looking at performance (financial, customer, internal business, and innovation and learning). The authors argue that financial performance, the drivers of it (customer and internal operational performance) and the drivers of ongoing improvement and future performance should be given equal

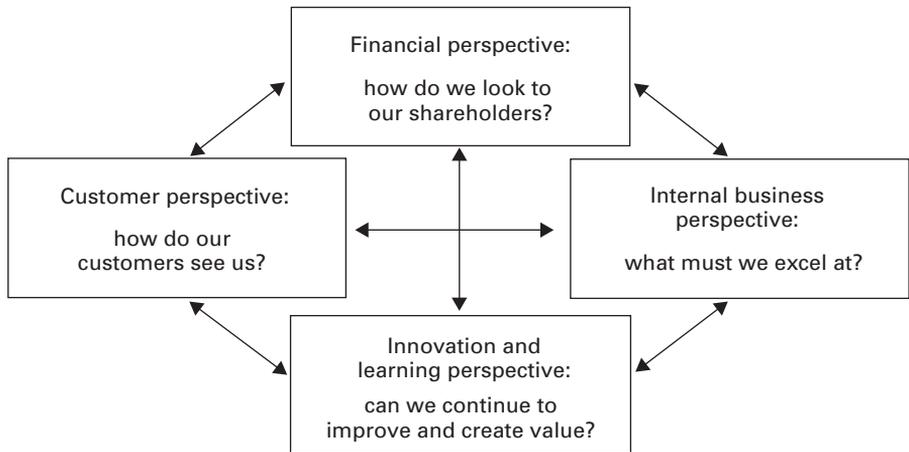


Figure 7.5: The balanced scorecard

Source: Kaplan and Norton (1992).

weighting (see figure 7.5). The Balanced Scorecard reflects many of the attributes of other measurement frameworks but links measurement to the organization's strategy more explicitly, especially in Kaplan and Norton's more recent writings (see, for example, Kaplan and Norton, 2004, 2006).

Kaplan and Norton argue that the full potential of the balanced scorecard will be realized only if an organization links its measures clearly identifying the drivers of performance (Kaplan and Norton, 1996b). Conceptually, this use of the scorecard is similar to the use of the *tableau de bord* (Epstein and Manzoni, 1997). Developed in France in the early twentieth century, the *tableau de bord* establishes a hierarchy of interrelated measures and cascading measures to different organizational levels, forcing the functions and divisions of an organization to position themselves in the context of the company's overall strategy.

Despite its widespread use, numerous authors have identified shortcomings with the balanced scorecard. It does not consider a number of features of earlier frameworks. The absence of a competitiveness dimension, as included in Fitzgerald *et al.*'s (1991) results and determinants framework, is noted by Neely, Gregosy and Platts (1995). In its traditional form, the question underpinning the customer perspective is "How do our customers see us?". No mention is made of the organization's performance relative to its competitors, other than implicitly – namely that the view any organization's customers have of it will be influenced by any experiences they might have had with that organization's competitors. Others critics

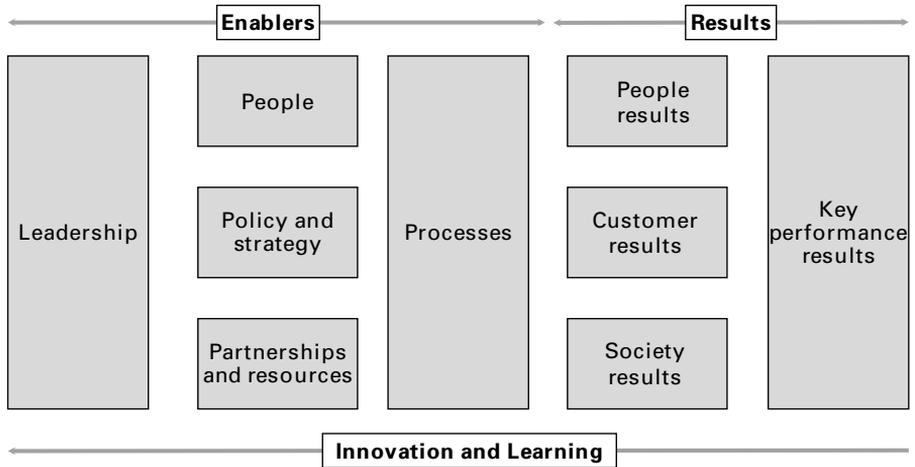


Figure 7.6: The European Foundation for Quality Management framework

comment that the balanced scorecard omits perspectives on human resources and employee satisfaction, supplier performance, product/service quality and environmental/community considerations (Maisel, 1992; Ewing and Lundahl, 1996; Lingle and Schiemann, 1996; Brown, 1996).

Although not designed as performance measurement frameworks, the European Foundation for Quality Management's Business Excellence Model and its US equivalent, the Malcolm Baldrige National Quality Award, take a broader view of performance, addressing many of the areas of performance not considered by the balanced scorecard (see figure 7.6). The Business Excellence Model is a broad management model that explicitly highlights the enablers of performance improvement and indicates result areas that should be measured. However, it is a self-assessment rather than an objective measurement framework, and the categories for measurement are very broad. Whilst the results areas are readily measurable, some of the enablers are not (Neely and Adams, 2001).

Characteristics of performance measurement frameworks

The performance measurement frameworks discussed in the previous section display a number of key characteristics that help an organization to identify an appropriate set of measures to assess their performance.

- (1) The work of Kaplan and Norton (1992) and Keegan, Eiler and Jones (1989) emphasizes the fact that the set of measures used by an

organization has to *provide a “balanced” picture of the business*. The set of measures should reflect financial and non-financial measures, internal and external measures, and efficiency and effectiveness measures.

- (2) The populated framework of measures should provide a *succinct overview of the organization’s performance*. For example, the simplicity and intuitive logic of the balanced scorecard has been a major contributor to its widespread adoption, as it is readily understood by users and easily applied to their organizations.
- (3) Each framework demonstrates the need for organizations to implement a set of performance measures that is *multidimensional*. This reflects the need to measure all the areas of performance that are important to the organization’s success. However, there is no consensus over what the dimensions of performance are. The EFQM model provides the broadest indication of dimensions of performance to be measured.
- (4) The performance measurement matrix provides *comprehensiveness*. It is possible to map all possible measures of an organization’s performance onto the framework and identify where there are omissions or where there is a need for greater focus. However, the PMM provides little indication of the different dimensions of performance that should be measured.
- (5) The tableau de bord and the SMART pyramid, along with the work of Bititci *et al.* (1998), explicitly demonstrate the fact that performance measures should be *integrated both across the organization’s functions and through its hierarchy*, encouraging congruence of goals and actions.
- (6) The tableau de bord and the work of Fitzgerald *et al.* (1991) explicitly, and the balanced scorecard and performance pyramid implicitly, explain how *results are a function of determinants*. This demonstrates the need to measure results and their drivers, so that the performance measurement system can provide data for monitoring past performance as well as planning future performance. In essence, measures can contribute both to an organization’s planning (feed-forward) and its control (feedback) system (Ballantine and Brignall, 1994).

As well as defining a similar set of “core” criteria, in their review of performance measurement frameworks, Ballantine and Brignall (1994) identify the need for a set of performance measures to reflect what they refer to as “non-core elements” of performance measurement systems. This indicates the need for the performance measures implemented to be consistent with management techniques and improvement initiatives that exist within the organization, such as benchmarking, activity-based costing management,

total quality management and business process redesign. There are other measurement frameworks and methodologies as well, such as shareholder value added or cost of quality, that have been developed to focus on the measurement of a specific performance issue (Neely and Adams, 2001). A comprehensive and multidimensional measurement system should encompass these measurement techniques.

So far, this contribution has presented the key attributes of existing performance measurement frameworks that enable them to help organizations identify the set of performance measures that appropriately reflects their performance and objectives. It is clear from the discussion that each of the frameworks presented in the literature falls short of satisfying these criteria in a number of areas. The remainder of the chapter presents a multifaceted framework, the performance prism, which attempts to address the shortcomings of the frameworks that are currently available, while satisfying the key criteria identified.

The performance prism

According to Teddy Wivel, senior partner in the Danish arm of Ernst and Young, “It will not be possible to create shareholder value without creating stakeholder value” (Crowe, 1999). Since the publication of Freeman’s book in 1984, considerable attention has been paid to the stakeholder approach to the management of organizations. In its *Tomorrow’s Company* report, the Royal Society for the encouragement of Arts, Manufacture and Commerce suggested that competitive success in the future will increasingly depend on taking an inclusive approach to management, reflecting the need for consideration of the requirements of all stakeholders to be central to performance measurement and management activities (RSA, 1995).

To reflect the growing importance of satisfying stakeholder requirements, the performance prism adopts a stakeholder-centric view of performance measurement. It is important to note that the performance prism does not assume that all stakeholders are equally important. Indeed, it is clear that in many organizations the shareholders will remain the most important stakeholders. Consideration should be given, however, to other important stakeholder groups, such as customers, employees and suppliers, all of whom are incorporated into the balanced scorecard, or variants of it.

In addition to these traditional stakeholders, the performance prism also considers groups of stakeholders that are growing in power and significance

in the current business environment, not least regulators, legislators and pressure groups. A key consideration for many organizations is how to satisfy their regulatory and legal communities. The regulators of the United Kingdom's recently privatized utilities have particular significance, for they often have the power to impose price restrictions, insist on operational investments and even revoke an organization's licence to operate if performance does not meet their requirements. Regulators are not confined to newly privatized industries, however. There are a variety of regulatory and legislative bodies seeking to prevent organizations from exploiting their competitive position, exploiting their employees or damaging the environment, for example. Regulators often provide a voice for stakeholders that do not have a collective voice, whilst pressure groups often express collective opinions and can have a significant influence on the operations of an organization.

In relation to stakeholders, the performance prism makes an important distinction between stakeholder satisfaction – what the stakeholders want of the organization – and stakeholder contribution – what the stakeholders contribute to the organization. This distinction, which is not made in other measurement frameworks, is important, because it highlights the fact that organizations have relationships with their stakeholders that involve mutual exchanges. Take, for example, customers. In recent years a growing number of authors have suggested that organizations should shift their focus from customer service and satisfaction to customer loyalty and profitability (Gittomer, 1998). All the same, this call to arms is a naïve one. Customers do not care whether or not they are loyal or profitable to the organizations that serve them. What they care about is the customer experience and the quality of service they receive. If this is appropriate, then they will reward the organization with their loyalty.

The same argument can be made for any of the stakeholders. Employees work for organizations in return for wages, opportunities to develop, and a rich and stimulating environment. Organizations employ people because they need access to their intellect and talent; they want their employees to be motivated and to give their best. In the human resources and psychology literature, our understanding of this bargain between employer and employee is becoming much more sophisticated through work on the psychological contract (Rousseau, 1995). This work needs to be extended to other arenas, however, in particular to studies of the psychological contracts between organization and customer, organization and supplier, and organization and regulator.

Once the questions about stakeholder satisfaction – “Who are the stakeholders and what do they want and need?” – and stakeholder contribution – “What does the organization want and need from its stakeholders?” – have been addressed, then it is possible to turn to the second perspective on performance: strategies. The key question underlying this perspective is: what strategies should the organization adopt to ensure that the wants and needs of its stakeholders are satisfied? In this context, the role of measurement is fourfold. First, measures are required so that managers can track whether or not the strategies they have chosen are actually being implemented. Second, measures can be used to communicate these strategies within the organization. Third, measures can be applied to encourage and incentivize the implementation of strategy. Fourth, once available, the measurement data can be analysed and used to challenge whether the strategies are working as planned (and, if not, why not).

The old adages “You get what you measure” and “You get what you inspect, not what you expect” contain an important message. People in organizations respond to measures. Horror stories abound of how individuals and teams appear to be performing well, yet are actually damaging the business. When telesales staff are monitored on the length of time it takes for them to deal with customer calls, it is not uncommon to find them cutting people off mid-call, just so the data suggest that they have dealt with the call within sixty seconds. Malevolently or not, employees will tend towards adopting “gaming tactics” in order to achieve the target performance levels they have been set. Measures send people messages about what matters and how they should behave. When the measures are consistent with the organization’s strategies, they encourage behaviours that are consistent with strategy. The right measures, then, offer not only a means of tracking whether strategy is being implemented but also a means of communicating strategy and encouraging implementation.

Many of the existing measurement frameworks and methodologies appear to stop at this point. Once the strategies have been identified and the right measures established it is assumed that everything will be fine. Yet studies suggest that some 90 per cent of managers fail to implement and deliver their organization’s strategies. Why? There are multiple reasons, but a key one is that strategies also contain inherent assumptions about the drivers of improved business performance. Clearly, if the assumptions are false, then the expected benefits will not be achieved. Without the critical data to enable these assumptions to be challenged, strategy formulation (and revision) is largely predicated on “gut feel” and management theory. Measurement data and its analysis will never replace executive intuition, but it can be used

to enhance greatly the making of judgements and decisions. One key judgement, of course, is whether an organization's strategy and business model remains valid.

A second key reason for strategic failure is that the organization's processes are not aligned with its strategies, and, even if its processes are aligned, then the capabilities required to operate these processes are not. Hence, the next two perspectives on performance are the processes and capabilities perspectives. In turn, these require the following questions to be addressed: "What processes do we need to put in place to allow the strategies to be executed?" and "What capabilities do/shall we require to operate these processes – both now and in the future?"

Again, measurement plays a crucial role by allowing managers to track whether or not the right processes and capabilities are in place, to communicate which processes and capabilities matter, and to encourage people within the organization to maintain or proactively nurture these processes and capabilities as appropriate. This may involve gaining an understanding of which particular business processes and capabilities must be competitively distinctive ("winners") and which merely need to be improved or maintained at industry-standard levels ("qualifiers").

Business processes received a good deal of attention in the 1990s with the advent of business process re-engineering. Business processes run horizontally across an enterprise's functional organization until they reach the ultimate recipient of the product or service offered – the customer. Michael Hammer, the re-engineering guru, advocates measuring processes from the customer's point of view; the customer wants it fast, right, cheap and easy (to do business with). Is it really as simple as that, however? Often there are many stages in a process. If the final output is slow, wrong, expensive and unfriendly, how will we know which component(s) of the process is/are letting it down? What needs to be improved? In the quest for data (and accountability), it is easy to end up measuring everything that moves, but learning little about what is important. That is one reason why processes need owners – to decide what measures are important, which metrics apply and how frequently they are to be measured, and by whom – so that judgements can be made upon analysis of the data and actions taken.

Processes do not function independently. Even the most brilliantly designed process needs people with certain skills, some policies and procedures regarding the way things are done, some physical infrastructure for it to happen and, more than likely, some technology to enable or enhance it.

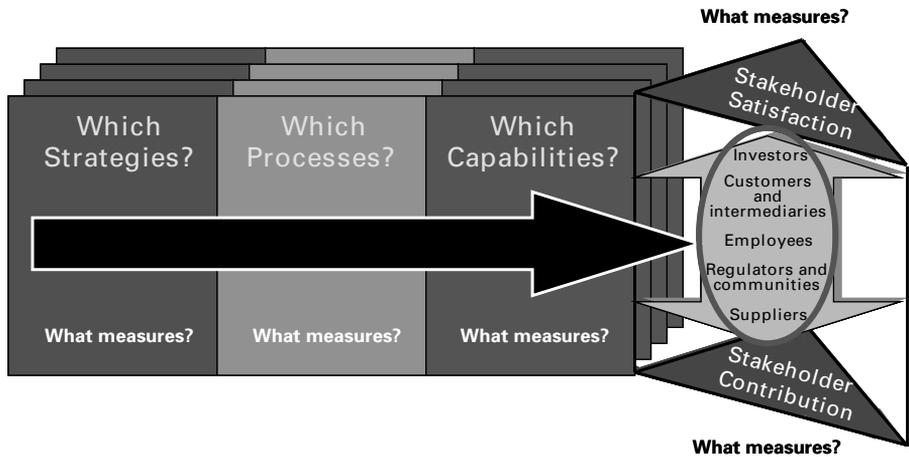


Figure 7.7: The performance prism

In fact, capabilities can be defined as the combination of an organization's people, practices, technology and infrastructure that, collectively, represents that organization's ability to create value for its stakeholders through a distinct part of its operations. Very often that distinct part will be a business process, but it could also be a brand, a product/service or an organizational element. Measurement will need to focus on those critical component elements that make it distinctive and also allow it to remain distinctive in the future. Competitive benchmarks will be needed in order to understand the size of the gap. Competitors will be seeking ways to create value for what is, probably, not exactly the same, but a very similar, set of stakeholders, too.

These five distinct but linked perspectives of performance are the basis of the performance prism framework (see figure 7.7). In essence, the performance prism identifies five questions for organizations to address when defining a set of performance measures.

- (1) *Stakeholder satisfaction* – who are our key stakeholders and what do they want and need?
- (2) *Strategies* – what strategies do we have to put in place to satisfy the wants and needs of these key stakeholders?
- (3) *Processes* – what critical processes do we need to operate and enhance these processes?
- (4) *Capabilities* – what capabilities do we need to operate and enhance these processes?
- (5) *Stakeholder contribution* – what contributions do we require from our stakeholders if we are to maintain and develop these capabilities?

Answering these questions forces those involved in designing the measurement system to articulate how their organization will create stakeholder value. In the performance prism, *results* (stakeholder satisfaction) *are a function of determinants* (the other prism, facets). Answering the questions at an organizational level provides a *succinct overview of the organization's performance*, but with considerable additional detail for each of the facets. These ensure that the framework is *comprehensive*, enabling all measures to be mapped on to it so that gaps in measurement can be identified. Consideration of each of the performance prism's facets ensures that the framework can be used at any organizational level, *integrated both across the organization's functions and through its hierarchy*. The authors consider the framework to be *multidimensional*, reflecting all the areas of performance that influence the performance of an organization. This enables a balanced picture of the business to be provided, highlighting external (stakeholder) and internal (strategy, process and capability) measures, as well as enabling financial and non-financial measures and measures of efficiency and effectiveness throughout the organization.

Lessons from practice

Since it was first introduced, the performance prism has been applied in a variety of organizations, encompassing various different settings. In this section we review the outcomes of these applications, particularly from the perspective of what can be learned about the performance prism in light of experience. For the purposes of this analysis we draw on six separate applications of the performance prism, ranging from the private to the public sector, in the United Kingdom, North America and the Middle East. We have chosen this diverse sample of organizations deliberately as we want to compare and contrast the application of the performance prism in very different environments. Table 7.1 summarizes the six applications.

Case A: DHL UK

DHL UK was one of the first firms to adopt the performance prism, and it was used during the development of the framework as a beta test site. In essence, DHL had decided that they needed a new way of conceptualizing performance in the organization. The business had tried to adopt the

Table 7.1: Case examples of organizations using the performance prism

Case	Description of organization	People involved in the implementation of the performance prism
A	DHL UK – 4,000 employees, 50 locations, \$500 million sales	Implementation facilitated by authors and developers of the concept
B	UK charity – 150 members helping 40 charities	Implementation facilitated by authors and developers of the concept
C	UK public sector – health care	Implementation facilitated by authors and developers of the concept
D	US public sector – employment services	Implementation facilitated by independent consultants
E	Canadian private sector – 7,500 employees, \$3.5 billion sales	Implementation facilitated by internal management on the basis of a book
F	Middle Eastern private sector – \$130 million sales	Implementation facilitated by internal management on the basis of a book

balanced scorecard previously, but it had not taken hold, partly because of the facilitation process used by the consultants. To DHL, the appeal of the performance prism was that it offered a broader framework for thinking about organizational performance, without being too prescriptive.

For DHL, the performance prism was used as a framework to stimulate thinking amongst the board members. All members of the board participated in brainstorming sessions designed to identify the wants and needs of the organization's key stakeholders, the strategies associated with delivering value to these stakeholders and the underlying processes and capabilities. An important development in DHL was the decision by the board to focus on questions rather than performance measures. Once the brainstorming sessions had been completed the board of DHL reviewed the results and decided what questions they, as a board, needed to be able to answer at their quarterly performance review meetings. Only when they had debated this did they turn to the question of what data (and hence what measures) might be needed to enable them to answer these questions.

DHL spent some considerable time building an infrastructure to support their performance management process. The business had already appointed performance analysts, but their role was changed following the implementation of the performance prism, and they were asked to focus on gathering and analysing the performance data needed to answer the key questions posed by the board. The principles of the performance prism were

also cascaded across the business, so that the performance review meetings at the regional and branch levels were aligned with the questions of strategic importance to the business.

Case B: a UK charity

The second case study involved a small UK charity that is effectively a charity body, providing services and support to other charities. Prior to the introduction of the performance prism the charities' performance measurement systems had been extremely rudimentary, and the management team decided to use the performance prism to help them think through appropriate measures for the charity.

Once again, a series of meetings were held, involving members of the management team and an external facilitator. The management team developed a success map for the organization based on the performance prism, and identified appropriate measures. These measures were designed using the performance measures design template.¹

At the implementation stage the project stalled, because the charity could not access the data necessary to make the measurements. Two years later, when we revisited the charity, they had successfully overcome these challenges and adopted a measurement system focusing on social impact.

Case C: UK public sector – health care

The third case involved a UK public sector health care organization. On this occasion the design and deployment of the performance prism was facilitated by two external facilitators, with the active participation of the organization's board. As with case B, the board developed a success map for the organization, chose appropriate performance measures in light of the success map and then designed the measures using the measures design template.

This organization devoted significant time and effort to the design and deployment of the performance measurement system, during the course of which the organization was able to "convert some the cynics". In essence, the performance prism formed a major part of an organizational culture change process that the chief executive was trying to drive. This change process survived, as did the performance prism, despite the fact that the chief executive left the organization midway through the deployment process.

¹ For a fuller description of this process, see Neely, Adams and Kennerley (2002).

Case D: US public sector – employment services

Unlike cases A to C, the fourth case did not involve the original developers of the performance prism, who found out about the application of their system only after the organization had used it and documented some of their experiences. In this case the process was led by independent consultants, who wanted a new framework to provide a structure for the board's annual retreat. They chose to use the performance prism for two reasons: because of its focus on a variety of stakeholders, and because of the concept of success maps.

The consultants used the performance prism to structure the discussion at the board's annual retreat, during which some significant strategic decisions were made. Following the retreat, the organization did not implement a performance measurement system but, instead, changed its operating model in light of the discussions that had taken place.

Case E: Canadian private sector

In the fifth case, the performance prism was introduced to the firm by an internal consultant, with no external support. The performance prism appealed to the consultant because of its focus on stakeholders and the distinction it made between stakeholder satisfaction and stakeholder contribution.

In this case the organization was seeking to redesign its business processes and associated performance measures. In the end the organization decided not to adopt the performance prism formally, but acknowledged that the framework had helped influence its thinking.

Case F: Middle Eastern private sector

In the final case, the performance prism was introduced to the firm by the chief executive, a UK national who had been appointed to run a business based in the Middle East. The chief executive was actively involved in all stages of the implementation of the performance prism, running workshops for his senior managers.

When he had joined the firm the chief executive had introduced a new vision and values programme, called Winning Hearts. While this had been successful, the chief executive felt that the programme lacked the "hard edge" provided by targets and measures, and hence decided to integrate the performance prism with the Winning Hearts programme.

This organizational development process continues today, with recent efforts to link individual performance assessment and appraisal to Winning Hearts and the performance prism.

The common themes

What are some of the common themes in this diverse set of applications? The first point to note is the variation in rationales for adopting the performance prism. Rarely do organizations say that they need a performance measurement system. Instead, they have some other pressing issue that they think a performance measurement system might help them address. In DHL's case the challenge was how to simplify, rationalize and align the organization's performance review processes. In case F the question was how to add a "hard edge" to a programme concerned with vision and values, while in cases D and E the question was how to stimulate senior managers to think more broadly about their responsibilities to their stakeholders.

The second point is the timescale associated with these projects. In the six case studies described the projects ran for periods of eighteen to twenty-four months, and in some cases longer. Designing and deploying a performance measurement system is not an activity that is completed quickly. Indeed, one could question whether it is ever completed. In each organization the solution to its question evolved during the course of the project, and continues to evolve today. Expecting an organization to adopt a measurement framework, crisply design and deploy it and then move on to another challenge would be naïve. Inevitably, measurement systems evolve in organizations, and they need constant attention.

The third point is the seniority of those involved in the projects. In each case the board were actively involved in the design and deployment of the performance prism, or associated system. There are two reasons for this. First, although described as measurement projects, measurement systems are intertwined with culture change and strategy deployment in organizations, both of which can preoccupy boards of companies. Second, as mentioned above, measurement projects are rarely short-lived projects. Indeed, they appear to evolve constantly. Such projects require the continued support of those who have the authority to allocate resources. Without this support, the projects are unlikely ever to gain traction.

References

- Ballantine, J., and Brignall, S. (1994). *A Taxonomy of Performance Measurement Frameworks*, Research Paper no. 135. Warwick: Warwick Business School.
- Banks, R.L., and Wheelwright, S.C. (1979). Operations versus strategy: trading tomorrow for today. *Harvard Business Review*, 57(3), 112–20.
- Bititci, U.S., Carrie, A.S., Turner, T., and Lutz, S. (1998). Integrated performance measurement systems: implementation case studies. In Bititci, U.S., and Carrie, A.S. (eds.), *Proceedings of the International Conference of the Manufacturing Value Chain on Strategic Management of the Manufacturing Value Chain*, 177–84. Dordrecht: Kluwer Academic.
- Brown, M.G. (1996). *Keeping Score: Using the Right Metrics to Drive World-Class Performance*. New York: Quality Resources.
- Bruns, W.J., Jr. (1998). Profit as a performance measure: powerful concept, insufficient measure. Paper presented at First International Conference on Performance Measurement, “Performance measurement – theory and practice”, Cambridge, 14–17 July.
- Chandler, A.D. (1977). *The Visible Hand: Managerial Revolution in American Business*. Boston: Harvard University Press.
- Crowe, R. (1999). Winning with integrity. *Guardian*, 27 November.
- Eccles, R.G. (1991). The performance measurement manifesto. *Harvard Business Review*, 69(1), 131–7.
- Epstein, M.J., and Manzoni, J.F. (1997). The balanced scorecard and tableau de bord: translating strategy into action. *Management Accounting*, 79(2), 28–36.
- Ewing, P., and Lundahl, L. (1996). The balanced scorecard at ABB Sweden: the EVITA Project. Paper presented at International workshop on cost management, Venice, 27–9 June.
- Fitzgerald, L., Johnston, R., Brignall, T.J., Silvestro, R., and Voss, C. (1991). *Performance Measurement in Service Businesses*. London: Chartered Institute of Management Accountants.
- Freeman, R.E. (1984). *Strategic Management: A Stakeholder Approach*. Boston: Pitman.
- Frigo, M.L., and Krumwiede, K.R. (1999). Balanced scorecards: a rising trend in strategic performance measurement. *Journal of Strategic Performance Measurement*, 3(1), 42–8.
- Gitomer, J. (1998). *Customer Satisfaction is Worthless: Customer Loyalty is Priceless*. London: Capstone.
- Hayes, R.H., and Abernathy, W.J. (1980). Managing our way to economic decline. *Harvard Business Review*, 58(4), 67–77.
- Johnson, H.T. (1992). *Relevance Regained: From Top-down Control to Bottom-up Empowerment*. New York: Free Press.
- Johnson, H.T., and Kaplan, R.S. (1987). *Relevance Lost: The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.
- Kaplan, R.S. (1984). Yesterday’s accounting undermines production. *Harvard Business Review*, 62(1), 95–101.
- Kaplan, R.S., and Norton, D.P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.

- (1996a). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- (1996b). Linking the balanced scorecard to strategy. *California Management Review*, 39(1), 53–79.
- (2004). *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Boston: Harvard Business School Press.
- (2006). *Alignment: How to Apply the Balanced Scorecard to Corporate Strategy*. Boston: Harvard Business School Press.
- Keegan, D.P., Eiler, R.G., and Jones, C.R. (1989). Are your performance measures obsolete? *Management Accounting*, 70(12), 45–50.
- Lingle, J.H., and Schiemann, W.A. (1996). From balanced scorecard to strategy gauge: is measurement worth it? *Management Review*, 85(3), 56–62.
- Lynch, R.L., and Cross, K.F. (1991). *Measure Up: The Essential Guide to Measuring Business Performance*. London: Mandarin.
- Maisel, L.S. (1992). Performance measurement: the balanced scorecard approach. *Journal of Cost Management*, 5(2), 47–52.
- Marr, B., Neely, A.D., France, M., Wilcox, M., Adams, C., and Mason, S. (2004). Business performance measurements: what is the state of the art? Paper presented at fourth Performance Measurement Association conference, Edinburgh, 28–30 July.
- Meyer, M.W., and Gupta, V. (1994). The performance paradox. In Staw, B.M., and Cummings, L.L. (eds.), *Research in Organizational Behavior*, 309–69. Greenwich, CT: JAI Press.
- Neely, A.D., and Adams, C. (2001). Perspectives on performance: the performance prism. *Journal of Cost Management*, 15(1), 7–15.
- Neely, A.D., Gregory, M.J., and Platts, K.W. (1995). Performance measurement system design: a literature review and research agenda. *International Journal of Operations and Production Management*, 15(4), 80–116.
- Neely, A.D., Adams, C., and Kennerley, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Stakeholder Relationships*. London: Financial Times/Prentice Hall.
- Neely, A.D., Bourne, M.C.S., Mills, J.F., Platts, K.W., and Richards, A.H. (2002). *Getting the Measure of your Business*. Cambridge University Press, Cambridge.
- Ridgway, V.F. (1956). Dysfunctional consequences of performance measurement. *Administrative Science Quarterly*, 1(2), 240–7.
- Rigby, D. (2001). Management tools and techniques: a survey. *California Management Review*, 43(2), 139–60.
- Rousseau, D.M. (1995). *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*. London: Sage.
- RSA (1995). *Tomorrow's Company: The Role of Business in a Changing World*. London: Royal Society for the Encouragement of Arts, Manufacture and Commerce.
- Silk, S. (1998). Automating the balanced scorecard. *Management Accounting*, 79(11), 38–44.
- Speckbacher, G., Bischof, J., and Pfeiffer, T. (2003). A descriptive analysis on the implementation of balanced scorecards in German-speaking countries. *Management Accounting Research*, 14, 361–87.
- Williams, S.M. (2001). Is intellectual capital performance and disclosure practices related? *Journal of Intellectual Capital*, 2(3), 192–203.

8

Beyond budgeting to the adaptive organization

Jeremy Hope

Introduction

Most performance management systems involve plans, targets and resources that are *negotiated, annual and fixed*. These systems were designed for stable trading environments in which suppliers could dictate to the market. As we know, however, markets are now unstable and customers rule the roost. Just think of the airline industry in recent years.

Most organizations expend huge amounts of energy (and money) trying to align and realign their strategies, structures and systems to meet this changing business climate. It is a difficult and expensive process and a nightmare for many employees, whose jobs are at risk. This inability to cope with discontinuous change is caused by management systems. Consequently, many leading-edge organizations are undergoing a historic shift from “make-and-sell” to “sense-and-respond” management. Whereas make-and-sell is an industrial-age model centred on transactions, capital assets, mass production, economies of scale and product margins, sense-and-respond is an information- and service-age model emphasizing client relationships, intellectual assets, mass customisation, economies of scope and value creation.

Toyota is a well-known example of a sense-and-respond organization. It operates with what is called a “pull” system. Instead of pushing products through rigid processes to meet sales targets, its operating systems *start from the customer; it is the customer order that drives operating processes and the work that people do*. The point is that, in sense-and-respond companies, fixed targets and rigid annual plans are anathema and represent insurmountable barriers. *That’s why adaptive organizations such as Toyota don’t have them.*

Systems based on sense-and-respond provide front-line managers with the capability to make fast and well-informed decisions. They enable divisional managers to see trends, patterns and “breaks in the curve” long before their competitors, and thus make crucial decisions regarding products and markets. They also enable senior executives to keep asking important

questions concerning strategic assumptions and risks while ensuring that operating units remain within acceptable performance parameters. Perhaps, above all, they help to share knowledge throughout the company, so that strategic changes can be made quickly and potential problems solved before they fester and grow.

However, none of this will happen unless the chief financial officer and the finance team replace their worn-out make-and-sell systems with those that support the sense-and-respond model. In particular, the CFO needs to implement practices and systems that enable continuous planning and decision making.

This chapter examines how organizations can adapt to emerging threats and opportunities using new, adaptive steering mechanisms based around a continuous planning cycle. It begins with a brief explanation of how “complex systems” make everything unpredictable. Then it suggests why CFOs need to:

- understand how complex systems make everything unpredictable;
- manage through continuous planning cycles;
- devolve planning to front-line teams;
- make rolling forecasts the primary management tool;
- report key metrics daily and weekly;
- manage resources dynamically;
- coordinate cross-company actions dynamically; and
- focus accountability on the relative performance of teams.

Understand how complex systems make everything unpredictable

We all rationalize reality in terms of *systems* that contain a set of interacting parts that create a result. Looking at the engine of a car, for example, enables us to understand how it moves. This Newtonian view of the world allows systems to be “run forward” in time and to provide predictions. However, since Darwin we have known that everything in the living world evolves and changes over time as it runs forward. This means that any predictions concerning the future of a particular system will be correct only if an evolutionary change does *not* occur. And this is *not predictable* from within the system itself. Systems are, clearly, more complex than we first thought.

When we delve further into “complex systems thinking” we begin to realize that a system can restructure and transform itself either as a response

to its changing environment or as a challenge to it. Nevertheless, at the lower level within a system, the “components” of a given system structure not only change their connections and interactions, but they can also change their own inner nature as they respond to local conditions. These changes can start small and rapidly escalate. The capacity of a system to evolve or transform will depend on the strength of its “defence” mechanisms that may be ready to spring into action and quell any incipient disturbance.

Seeing social systems such as organizations as complex systems has profound implications. Instead of viewing them as functional machines the performance of which can be optimized, they are in reality creative, adaptive entities that explore, experiment and learn over time, changing their goals and strategies, and transforming themselves and their environment. This means that, instead of basing our strategies and actions on *prediction*, with the development and implementation of a plan designed to take us from “here and now” to “there and then”, we have to adopt a more creative approach. This implies more frequent monitoring and reassessments, with an awareness and capacity to change course so as to make use of what works and reduce reliance on what doesn’t. This is an approach that recognizes the constant need to learn about what is happening and to try to make sense of it as quickly as possible.

Complex systems theory provides us with a decision framework that resonates with the real world for most people. They inherently know that decisions are fraught with risk, yet their business leaders demand certainty wrapped up in neat and tidy spreadsheet models with reams of numbers that justify why a particular project should go ahead. There is more judgement in all decisions than perhaps we realized. Toyota’s approach is to be as thorough as possible and minimize the risks by gathering all the current knowledge to provide a sound decision base. Every variable needs investigation. Numbers should be treated with caution. Human intelligence, intuition and experience are the key factors.

Devolve planning and decision making to front-line teams

The word “planning” is often used interchangeably with “strategy”. But “strategy” is one of those words (like “culture”) that is commonly used in the language of business but with little common understanding of what it means. There is, of course, no single meaning. Meaning depends on context. In 1987 Mintzberg described strategy in terms of “five Ps”: strategy as

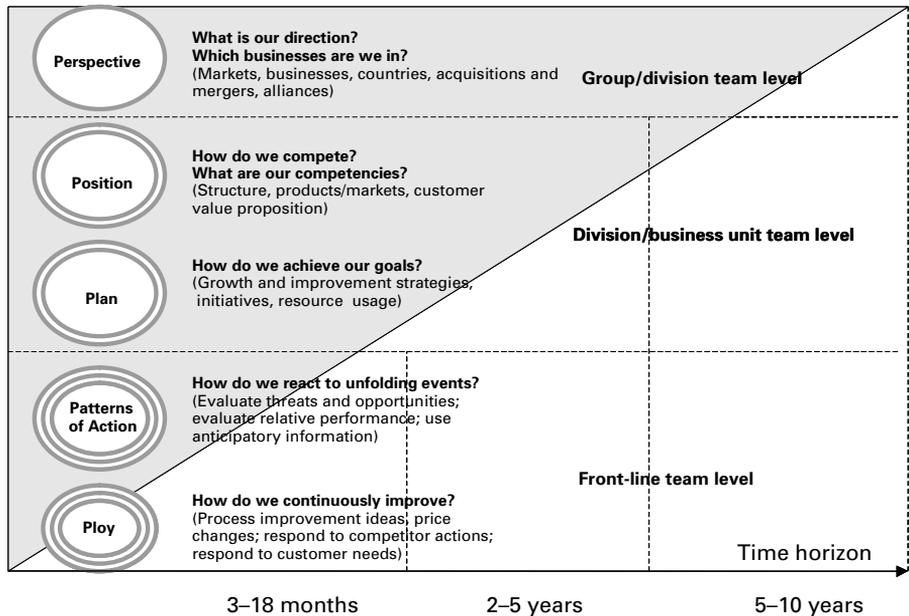


Figure 8.1: Mintzberg's different views of strategy

Source: Mintzberg (1987).

perspective, position, plan, patterns of action and ploy (see figure 8.1). These different views of strategy help us to understand the roles and time-based perspectives taken by teams at different levels of the organization. The number of ellipses at each level reflects the greater number of expected planning reviews as you get nearer to the front line. Group and large business division executives, for example, will take a long-term view considering such questions as “What is our direction?” and “Which businesses should we be in/not in?”. The business unit team will take a medium-term (two- to three-year view) and ask questions such as “How do we compete?”, “What are our competencies?” and “What should our goals be?”. Front-line teams will take a shorter, more tactical view, over a period of between three months and two years. They will address questions such as “How do we react to unfolding events?” and “What tactics should we use?”.

While the strategic direction at the group level and market positioning at the business-unit level need to be as stable as possible, front-line teams need to plan continuously how to improve their performance and respond rapidly to changing events (where an organization draws the front line will depend on a host of factors, including competencies and information availability). Thus, continuity of strategy and rapid change are not inconsistent. In fact,

“the two go right together”, notes strategy expert Michael Porter, “if we make the very important distinction between strategic positioning and operational effectiveness in executing that position” (Porter, 1997). According to Porter,

Positioning is a strategy in which the goal is to be different, and in which trade-offs are necessary between what a company is doing and what its competitors are doing. In order to create those conditions, however, one needs continuity in the basic positioning – the types of products it offers, the essential core of the value it is delivering. But within that continuity should be a feverish and on-going process of change. If one examines high performance companies, and by that I mean those that regularly outperform their industries, one finds that they don’t change their strategies. Instead of change, we see stability. We see continuity. But we also see a lot of change in the details – the product details, the service details.

In today’s fast-paced information economy the people best placed to see the changes taking place within markets, competitors and customers are front-line teams. They are the ones who have to respond quickly to emerging events, and this means devolving to them the power and responsibility to take the necessary action. Nor can this process be deterministic and fixed. It should be driven by (internal or external) events rather than by the annual planning cycle. Planning needs to become a continuous cycle with new knowledge feeding the process. The degree of freedom depends on the maturity of the model. At Swedish bank Handelsbanken, branch managers have significant freedom, but this has gradually been extended over many years. Although the bank has five product-based subsidiaries and a central product development group, there are no product-based targets in the regions and branches. It is up to the product groups to “sell” their products through the branch organization. Though they have flexibility over prices, discounts and which products they sell, branch teams clearly work within acceptable limits.

Manage through continuous planning cycles

The typical approach to planning is based on a “take that hill” view of strategy. Planners, like military commanders, set an objective, formulate a plan, communicate it to their superiors and then apply all their resources to achieving it, irrespective of what might happen in the meantime. Leaders in adaptive organizations say this makes little sense in times of rapid and discontinuous change. Indeed, “the hill” may well have disappeared

from the landscape before the attack can be made. Only by probing and scanning for changes, they argue, can firms take advantage of fast-emerging opportunities and counter unanticipated threats. However, they also know that they need some stability in their strategic direction and positioning and lots of flexibility at the tactical level. They need capable and committed people at all levels to use their skill and intuition to make fast and effective decisions.

Adaptive organizations believe that discontinuous change is now the norm. They see planning as a continuous, inclusive process, driven by events (such as the launch of a new product or a competitive threat) and emerging knowledge, and not constrained by the current financial year. Mintzberg uses the term “natural continuous improvement” to describe this approach to strategy making. He also calls this management style “quiet management”, which he notes is “not about drinking champagne in business class; it is about rolling up sleeves and finding out what is going on. And that is not parachuted down on the organization; it rises up from the base” (Mintzberg, 1999).

Mintzberg captures the whole flavour of strategy making in adaptive organizations. It doesn’t need sophisticated tools. It relies on fast, relevant (actionable) information and responsible people who know what is expected of them and what to do in any given situation. The organization operates as a team, rather like an orchestra, with the conductor providing strategic direction, harmonizing and coordinating activities, and setting and upholding values and standards.

The emerging continuous, or “event-driven”, planning cycle has four steps: check, aim, plan and act.¹

- (1) *Check* – the starting point. Where are we right now? What does the short-term future look like?
- (2) *Aim* – the next step. Are we on a trajectory to meet our aspirational goals? Does our strategy need to change?
- (3) *Plan* – the third step. What actions do we need to take to improve our performance? What impact will these actions have on our performance?
- (4) *Act* – the fourth step. How should we execute the plans and manage the existing business?

¹ This sounds similar to Dr Deming’s “plan–do–check–act” (PDCA) cycle. In fact, it’s quite different. Deming was talking about a manufacturing system or subsystem, whereas I am referring to a business planning system. For example, when Deming talked about “plan” he meant having an idea for improving the system; check was “See if the idea works”, “do” was “Put it in the line” and “act” was “Go live”.

You will note that nowhere in this check–aim–plan–act cycle has the team made a commitment to a higher authority to reach a specific target. In other words, there is no fixed performance contract. All the commitment to improve is within the local team. This taps the power of intrinsic motivation. It is the team that sets the goals and plans and it is the team that has the drive to make them succeed. This *local* check–aim–plan–act cycle is typical of many adaptive organizations. However, the key to its success is that it is driven *locally* by people who want to improve their relative performance.

How does this cycle work? Imagine you have just been appointed as a branch manager at a retail bank. You have inherited a poor situation. First, you *check* where you are. You are at the bottom of the league table and your cost-to-income ratio is lagging behind. Second, you *aim* for an improvement goal. You consult with your team and your regional managers and you decide to set a goal to improve your cost-to-income ratio by 30 per cent over two years. Thirdly, you *plan* what to do. You discuss with your team the options available for achieving this goal. You then test the options by examining their likely forecast outcomes. You would also have a dialogue with your regional manager about the range of options available and ask his advice. He would ask questions about assumptions, risk and timescales, and may even provide some input about best practices elsewhere. Finally, you *act*. You choose those initiatives that offer the best chance of success. You then acquire (or reduce) the resources you need to put the plans into action. This cycle repeats itself endlessly as the branch continuously tries to improve its performance.

Make rolling forecasts the primary management tool

Forecasting is another key feature of the adaptive planning model. Most short-term forecasting processes, however, are invariably confined to asking the question “Are we on track to meeting our targets and, if not, what action do we need to take?”. Such forecasts are not aimed at supporting strategy, and what managers often find is that there’s a gap between October and January (as last year’s accounts and next year’s budgets are being finalized) when no one is looking at likely future performance. Adaptive organizations know that their operations don’t switch off on 31 December each year and start again on 1 January. They deal with these problems by moving to monthly or (more commonly) quarterly *rolling forecasts*.

Let's assume we are just approaching the end of the first quarter. The management team gets the rough figures for that quarter and starts to review the next four quarters ahead. Three of those quarters are already in the previous forecast, so they just need updating. A further quarter, however, needs to be added (Q1 for the next year). More time will be spent on the earlier quarters than the later ones, using as much relevant knowledge and business intelligence as can be gathered.

Leading organizations are placing forecasting at the centre of the management process. It becomes an essential tool for business managers to support their decision making, rather than another management chore that needs to be done to feed the corporate beast. Focusing on only a few key drivers, the process should take no longer than a day or two, and it should be done not by specialist finance people but by the business team itself.

Report key metrics daily and weekly

To be adaptive, managers must have effective “radar screens”. Key performance indicators (KPIs) provide managers with advanced warning signals that there are problems looming ahead and that action needs to be taken. However, there is much confusion about KPIs, what they mean and how they should be derived. In fact, the KPIs in most common use are *key result indicators* (KRIs) – a different type of measure. Whereas KPIs are *leading* performance measures, KRIs are *lagging* measures, which include customer satisfaction, net profit before tax, the profitability of customers, employee satisfaction and return on capital employed. The common characteristic of these measures is that they are the result of many inter-related actions. They are designed to tell you whether you are travelling in the right direction, but not whether you are travelling fast enough or you have enough fuel. They provide information that is ideal for the board (i.e. to those not involved in day-to-day management) but not for operating managers.

Separating KRIs from KPIs has a profound impact on reporting. Managers are able to distinguish between measures that impact high-level governance and those that impact operating management. In other words, an organization should have a governance report (ideally in a dashboard format) made up of five to eight measures providing high-level key result indicators for the board, and a balanced scorecard comprising up to fifteen measures for operating management.

There is a story about Lord King, who reportedly set about turning British Airways (BA) around in the 1980s by concentrating on just one KPI. He was notified, wherever he was in the world, if a BA plane was delayed. The senior BA officials at all the relevant airports knew that if a plane was delayed beyond a certain threshold they would receive a personal call from the chairman. It was not long before BA planes had a reputation for leaving on time. This KPI affected all the balanced scorecard perspectives. Late planes increased costs in many ways, including additional airport surcharges and the cost of accommodating passengers overnight as a result of planes being “curfewed” due to noise restrictions late at night. They were also a primary cause of dissatisfied customers and alienated those people meeting passengers at their destination (possible future customers). They used more fuel, as planes had to circle the airport when they missed their landing slots. And they would disrupt servicing schedules, resulting in poor service quality and increased employee dissatisfaction as they had to deal with customer complaints.

True KPIs should be reported and monitored on a daily or weekly basis. A quarterly or annual measure cannot be a KPI. For example, if customer satisfaction is measured no more than six-monthly or annually it cannot be an “actionable” driver. It is, therefore, at best only a KRI, not a KPI. A KPI should indicate what action needs to take place. The British Airways “late plane” KPI immediately illustrated the need to focus on recovering lost time. As a result, cleaners, caterers, ground crew, flight attendants, liaison officers and traffic controllers would all work together to save precious minutes. A good KPI has a knock-on effect. Improving quality, for example, lowers defect rates, speeds up inventory turns and, ultimately, increases profitability.

Manage resources dynamically

If business unit teams are expected to use their intuition and judgement to make quick decisions, they must have quick access to resources – especially to new operating capacity, including people and technology. At US airline Southwest Airlines, employees can acquire supplies to fix local problems. They don’t need to get authorization for purchase orders. Handelsbanken branch managers can hire staff according to demand. This was one of those “surprises” in the devolution process. When branches were given this discretion, sceptics thought that staff levels would rise, but the opposite

happened. Managers, conscious of their need to raise profits, were no longer inclined to hold onto staff when demand began to fall (on the assumption that it would be hard to regain approval to rehire people). Thus, staff levels and costs fell significantly. The lesson is that, when resources belong to “the company” and are negotiated through budgets, people feel little responsibility for them. If they are accessed by the local team as part of their action plan, however, ownership and responsibility become much more personal.

Changing direction with its old forecasting system was not easy at American Express. Each of its three operating segments had its own approach to forecasting, using simple spreadsheets based on its different markets (75 per cent of forecast data had to be manually keyed in). From these individually prepared forecasts, the company had to consolidate the data. Under the old system, it had seventy-five levels of review and consolidations. This involved a tremendous amount of labour and time to produce a forecast. Being able to do quarterly reviews and timely investment analysis under this system was impossible.

Now managers review performance monthly, and these reviews link to and inform forecast updates. They also include competitor analyses and reviews of changes in the external environment. They identify and quantify emerging risks and opportunities and the impact on plans and investment decisions. Investments are focused at the total-company level (as apposed to business-unit level) and funds are released dynamically according to current forecast and improvement priorities, instead of annually in advance. Monthly refreshes of key revenue and expense line item forecasts drive resource allocation decisions. Full forecasts are updated three times per year.

Resource decisions have moved to front-line teams, with the board (or an appointed subcommittee) in control of the strategic project portfolio and the prioritization of resources. This committee is constantly looking at rolling forecasts and releasing funds on the basis of capacity plans and strategic initiatives. This process tells the committee members what funds they have available, how many funds are already committed and what is left to release into the system. This approach has cut costs dramatically, as capacity is not fixed months or years in advance on the basis of unrealistic assumptions.

Organizations such as Toyota use twelve-monthly rolling forecasts (updated every month) to review capacity requirements. There is one forecast for each product family. The operations capacity is then established for each month, taking into account the revised product mix forecasts and any bottlenecks within the value stream. The operations team makes today what customers require today. Materials and components are pulled into the

system “just-in-time”, resulting in low inventories and perfect on-time delivery. To cope with this there must be spare capacity. Once the forecasts have been prepared the management team look at forecasts versus capacity. If there is further capacity required then they simply provide additional people. If there is not enough capacity then they need to think imaginatively how this can be made available. This can be done through further lean improvements, changes in the production process, changes to the manning of production cells, the acquisition of additional equipment or the outsourcing of some products. If the capacity problem is immediate then they would work overtime, or maybe get help from another value stream (Maskell and Baggaley, 2004, 191–2).

At Toyota, planning is a continuous process. By the early 1990s, while each plant forecast what its likely output (and thus capacity) would be, detailed planning was only one month ahead, with March’s planned production being set in the second half of February. As Toyota has now moved to a single common platform in nearly all its global plants (these can handle up to eight different models, never mind varieties of the same model!), it has even more planning flexibility and is able to reduce cycle times even further. The cycle time from order to finished product is moving towards a staggering five days (Brown, 2004). This additional speed and flexibility reduces costs further still, making it even more difficult for competitors to catch up.

Coordinate cross-company actions dynamically

In most organizations individual departments and business units fiercely protect their own parts of the business. The real enemies appear to be other business units rather than external competitors. Adaptive organizations remove these boundary walls and the processes (such as performance contracts) that reinforce them. They enable teams to coordinate their plans and actions dynamically in response to customer demand. They work to cut lead times and provide a seamless service to customers.

Adaptive organizations rely on fast feedback from internal customers to respond to market changes. At Handelsbanken, for example, central product development units develop new products based on customer reports rather than making their own assumptions about matching competitor offerings. Toyota changes cars on the line to meet exact customer needs. At Southwest Airlines, there are plenty of ways that ideas are shared and actions coordinated across the company. There is a Tuesday meeting that

examines capital spending and reviews business performance. Chairman Herb Kelleher believes that because leaders cannot possibly know everything that goes on in the organization – and don't pretend to – this is a source of competitive advantage. “The freedom, informality, and interplay that people enjoy,” he notes,

allow them to act in the best interests of the company. For instance, when our competitors began demanding tens of millions of dollars a year for us to use their travel agents' reservations systems, I said, forget it; we'll develop an electronic, ticketless system so travel agents won't have to hand-write Southwest tickets – and we won't be held hostage to our competitors' distribution systems. It turned out that people from several departments had already gotten together, anticipated such a contingency, and begun work on a system, unbeknownst to me or the rest of our officers. That kind of initiative is possible only when people know that our company's success rests with them, not with me. (Kelleher, 1997)

Focus accountability on the relative performance of teams

Implementing adaptive management systems will be ineffective unless senior managers take action to deal with target setting and performance evaluation. A number of organizations have answered this question by abandoning annual budgets and financial targets altogether. They now manage performance in a fundamentally different way. Goals and plans are devolved to local teams, who set medium-term (typically two- to five-year) “aspirational” goals aligned with the organization's strategic direction. A “rolling review” process then re-evaluates these medium-term goals every year. However, the most important change is that these goals are *not fixed for the year*. They do not (explicitly or implicitly) form a fixed performance contract. High expectations lead to tough standards that managers are expected to meet, but they are evaluated and rewarded *after the event* according to how they performed against those standards in the light of the circumstances that actually prevailed, and, perhaps more importantly, how they performed against their peers. Thus, there is still a “performance contract”, but it is based on some form of *relative result*. The benefits are that the process is fast and, because the benchmarking bar is always being raised, it is more likely to maximize profit potential.

Although managers know what constitutes “acceptable performance” (or the “benchmark expectation”), they generally don't know until the end of

the year how well they have performed and thus how much bonus they will receive. They also know, because performance is evaluated against a range of factors relative to competitors, the market and maybe the previous period, that they can achieve good bonuses in a low-profit year and poor bonuses in a high-profit year. The relevant assessment question is: “Did they do as well as they could have done, given what we now know about the profit-making opportunities during the period and what the competition has achieved?” This approach focuses managers on maximizing profits at all times rather than playing games with the numbers. This is because there are no fixed points that lead to irrational behaviour. The downside is that the level of the bonus is decided by peer reviews, which are subjective. However, most firms have used a formula to provide some structure to these reviews.

This approach to goal setting can be likened to a motor race in which each driver has to beat the competition while at the same time dealing with many unknown factors that will determine the outcome, such as the performance and behaviour of other drivers, the reliability of the car and the weather conditions. All the drivers know roughly what has to be done prior to the race to improve their performance and, preferably, win the race; but only with hindsight will they know how well they have performed. Firms that have gone “beyond budgeting” have used a variety of approaches to adopt this type of “relative performance contract”.

How teams are recognized and rewarded also needs to change. Table 8.1 shows the method used by Jean-Marie Descarpentries at French computer manufacturer Groupe Bull in the mid-1990s. Notice the mix of relative measures. They include not only growth and profit versus the competition but also growth and profit against the previous year.

This is how it worked. Maximum performance bonuses were set at 30–50 per cent of salary at the executive level, 20–30 per cent at the operating level and less (although not zero) in other areas. What was interesting, though, was how the actual payout was calculated. Take a business unit. If a business unit employee had a base salary of \$50,000 and the maximum bonus was set at 30 per cent (or \$15,000) and the formula set the payout at 60 per cent, then the final payout would be \$9,000. Both the corporate president and his executive committee would review performance independently. First they would examine growth versus last year and against the competition, bearing in mind all the competitive factors that pertained during the period. Secondly, they would examine profitability. Thirdly, they would look at debt. Finally, they would look at certain qualitative

Table 8.1: Performance appraisal formula for a business unit at Groupe Bull in the mid-1990s

Key performance indicator	Weighting	Total score	Weighted score
Growth versus previous year	20	50	10
Growth versus competition	20	40	8
Profit versus previous year	20	60	12
Profit versus competition	20	50	10
Debt versus previous year	10	80	8
Quality factors versus previous year	10	60	6
<i>Executive committee evaluation</i>			54 %

factors such as employee turnover. The choice of balance sheet and strategic measures might be different for each business unit. This assessment set the bonus levels of all managers and employees within that particular business unit. If a business unit team underperformed, they were given a second chance. If they underperform again the following year, however, they are likely to receive dismissal notices or be moved elsewhere.

Handelsbanken has elevated peer-based performance reviews to a whole performance culture based on competitive success. They both rely on setting high performance standards and then leaving local teams to work out how they will exceed them. Thus, performance rankings are a strong feature and are used for peer-to-peer comparisons at every level. In this culture, all the teams know where they are in relation to others, and thus must set their targets at a point that they believe will enable them to improve their position. They must be ambitious, because other teams are also trying to improve, so they must improve at an even faster rate. The additional benefit is that each team has a self-adjusting target-setting process geared to continuous improvement. The role of leaders is to challenge ambition and convince people that they can achieve extraordinary performance.

To support this approach it uses only one, group-wide, profit-sharing scheme for all employees. Thus, there are no “incentives” for any team or sales person to achieve a specific target. Nor are branches rewarded for their placement in the performance league table (although they are *recognized* in different ways, thus increasing the importance of league tables). They use the language of gain sharing (sharing in the fruits of collective success) as opposed to individual incentives (“You must achieve X result to earn Y bonus”) to provide people with a stake in the success of the organization. Handelsbanken executives believe that their group-wide profit-sharing scheme is an important element in removing the “cellular” or “defend your

own turf” mentality that pervades many organizations. It avoids rewards becoming “entitlements” that, if not received, can lead to a disaffected and, in some cases, a demoralized workforce. Since abandoning budgeting and adopting relative measures in the 1970s, Handelsbanken has outperformed its rivals on just about every measure you can think of, including return on equity, total shareholder return, earnings per share, cost-to-income ratio and customer satisfaction.

Moving to relative measures and performance rankings is not the perfect solution (achieving “like-for-like” comparisons, for example, can be difficult), but many firms believe that the benefits outweigh the problems. These include huge cost savings from not preparing and negotiating annual budgets, eradicating much of the “gaming” process (there is little point in gaming if there is no fixed target to meet), faster response from using more adaptive processes, better strategic alignment between goals, plans, measures and behaviour, and more value-adding work from the finance team. Moreover, they also believe that they recognize and reward the best performers, and not just those who are skilled at negotiating budgets or those who have been fortunate to hit a market on the upturn that wasn’t anticipated at the time of the budget agreement. In other words, basing performance evaluations on benchmarked KPIs (which have been well chosen) provides a fairer reflection of how well managers have done, as all the “competitors” face the same economic uncertainties. While “relative” benchmarks can, of course, go down as well as up, all firms set high standards that managers are expected to achieve.

Many executives find it difficult to imagine corporate life without fixed targets and annual budgets. For example, they worry about the reactions of analysts. If CEOs don’t provide such promises (or even “nods” and “winks”), how might analysts react? The answer is that they will make up the numbers themselves, but at least they will have to find out how the company is performing to reach their own conclusions. We have recently seen a number of companies, including UBS, Southwest Airlines, Nokia, Coca-Cola and Gillette, discontinuing the practice of giving fixed targets to analysts. It is also worth noting that Warren Buffet believes that fixed targets make little sense. It is no accident that he sits on the board of two of the companies just mentioned (Coca-Cola and Gillette) and that his own company, Berkshire Hathaway, also refuses to give targets.

At the management level, we are also seeing a disengagement from fixed targets. Many organizations are realizing that measuring against peers, market movements and benchmarks makes more sense and doesn’t need a

protracted period of negotiation. It is instructive to note that organizations such as Southwest Airlines, Handelsbanken and UK manufacturer Tomkins no longer use fixed performance contracts anywhere in their businesses. It is probably the single greatest disabler of adaptive, lean and ethical organizations in force today.

Investors want to back those companies that are consistently at (or near the top) of their peer groups. Dr Jan Wallander, architect of the Handelsbanken “beyond budgeting” model, believes that “*the fundamental purpose of a firm in a market economy is to deliver as high a return on capital invested in the company as possible. A company is successful and will survive if it gives a higher return than other companies in the same field. The real target is thus not an absolute sum in dollars and cents but a relative one. Beating the competition is the real target.*”

References

- Brown, S. F. (2004). Toyota’s global body shop. *Fortune Magazine*, 9 February, 68–70.
- Kelleher, H. (1997). A culture of commitment. *Leader to Leader*, 4, 20–4.
- Maskell, B., and Baggaley, B. (2004). *Practical Lean Accounting*. New York: Productivity Press.
- Mintzberg, H. (1987). Five Ps for strategy. *California Management Review*, 30(1), 11–24.
- (1999). Managing quietly. *Leader to Leader*, 12, 24–30.
- Porter, M. (1997). Creating tomorrow’s advantages. In Gibson, R. (ed.), *Rethinking the Future*, London: Nicholas Brealey.

Theoretical conditions for validity in accounting performance measurement

Hanne Nørreklit, Lennart Nørreklit and Falconer Mitchell

Introduction

The concept of truth is a foundation upon which the accounting profession has built its reputation. The need for truth in accounting is enshrined in ethical codes, accounting regulations, authoritative texts and, of course, the auditor's assignation of a "true and fair view". MacNeal, in his classic critique of accounting, has emphasized the salient role of truth for the accountant:

Accounting is the language of finance. Members of the accounting profession are interpreters upon whom the vast majority of people must rely for information . . . If interpreters do not tell the truth, or tell truths intermixed with half truths, many people will be deceived to their hurt. (MacNeal, 1970[1939], 1)

The absence of truth from accounting information, or, indeed, even doubts about its presence, can have profound effects upon the accounting profession, its operation and its regulation. For example, in the United Kingdom in the late 1960s and 1970s the criticism of the flexibility of accounting practice during the GEC/AEI merger and the Leasco Pergamon affair preceded and triggered the implementation of a national system of accounting standardization (Stamp and Marley, 1970; Leach and Stamp, 1981; Stamp 1985). More recently, in the United States, the Enron scandal, and subsequently the Worldcom scandal, destroyed a leading international firm of accountants – Arthur Andersen – and led to a marked strengthening of formal corporate governance regulations.

Truth is, therefore, a very practical and immediate issue for accountants. It is a vital key to the validity of the accountant's work, and the truthfulness of information is a crucial claim that underpins the credibility of accountants. In the post-Enron world the truth of accounting information, both financial

and managerial, has been put under the spotlight and has assumed even greater significance.

Truth for the accountant has traditionally been founded in the belief that, in reporting on performance, accountants provide a neutral and objective representation of an underlying reality (see Solomons, 1991a, for a lucid exposition of this position). Accountants do not deal in fiction or untruths. From this perspective, accountants can be considered to follow a paradigm of realism. Real-world events and objects, which can be translated into accounting information, provide substantiation for their outputs. However, it has also long been recognized by accountants themselves that the paradigm of realism is difficult to use as a comprehensive justification for their actions. Many aspects of accounting practice do not readily fit this paradigm. For example, the estimates and judgements inherent in depreciation, cost allocations for stock valuation, the establishment and adjustment of reserves and provisions and the treatment of goodwill do not sit well with the qualities of information neutrality, objectivity and conformance with reality. There are no real-world referents for these aspects of accounting (Thomas, 1969, 1974), and therefore the paradigm of realism breaks down.

Moreover, alternative paradigms can be applied to the accounting world (see, for example, Tinker, 1991). From a social constructivist perspective, it is assumed that human thoughts, discourse, agreements and concepts create the (accounting) world. Thus, inevitably, information produced by accountants is socially constructed. Consequently, the process of producing accounting information on this basis rules out the very qualities that accountants value and promote – i.e. neutrality, objectivity and realism. The ability to deliver truth is, therefore, compromised by the very nature of the human condition and circumstance. Truth, in the sense that accountants have traditionally conceived of it, does not exist. Accountants may want to act under the paradigm of realism and may say that they do but, in their discipline, this is impossible.

This chapter addresses this fundamental question: the extent to which accounting performance measurement can be considered to have a justifying base in truth. It does so by, first, outlining an alternative paradigm for accounting; second, reviewing different concepts of truth relating to this paradigm; third, developing a specific concept of truth that both fits this paradigm and that is applicable to accounting; and, finally, describing how this concept of truth can be used to create a stepped process for the practice of accounting performance measurement.

The concept of reality

An alternative paradigmatic position labelled “pragmatic constructivism” is proposed in this chapter (Nørreklit, Nørreklit and Israelsen, 2006). Ontologically, pragmatic constructivism takes its point of departure in the concept of reality. It suggests that the actor continuously constructs his relationship to the world. “Reality” is contrary to fiction, illusion, dreams and similar items. Since life is always life *in* a world, we consider reality as some kind of relation between the actor – whether a person, or even an organization – and the world, while fiction and illusion express a missing or faulty relation between the two. If this construction fails, then the understanding of reality that guides the actor is erroneous, in the sense that it leads to unsuccessful actions – i.e. activities with poor performance. Four dimensions of reality must be integrated in the actor–world relation if the construct is to be successful and open up reality to the actor. The four dimensions are facts, logic, values and communication. Maintaining this integration is, therefore, a main task of managing and controlling any acting unit, be it an organization, a firm, a family or a society.

Facts

Facts are based on a source that exists independently of the observing actor. They may or may not exist independently of any actor or any language. A tree exists independently, but the fact that it is a tree does not. Thus, the paradigm of pragmatic constructivism recognizes that there is a physical and biological world, which exists independently of human consciousness.

However, a thing does not exist as a fact until it is recognized and established as a fact. There is nothing we can call a “fact” unless it has been recognized and established as a fact by some actors. The factual is the converse of, for instance, dreams, wishful thinking and deception. However, something imagined – a dream, a wish or a lie – is itself an experienced fact. Thus, there are all kinds of facts, but whether what is imagined, dreamed, wished or told is a fact is a different matter. For instance, the €100 bill in my wallet is real because it is a fact that it is in my wallet, but the €1 million in my wallet is a dream simply because it is a fact that it is not in my wallet. To establish successful control, knowledge about relevant facts is inevitable.

Logic

Logic is a rational element inherent in our ability to calculate and reason in a stringent, logical manner. Mathematics, statistics, formal logic and the like are recognized disciplines in logic and play a certain role in the methodology of scholarly work. However, logic also concerns reasoning through the use of concepts and the building of concepts, including ones that are essential to our lives. We use logical operations to reflect and recognize logical possibilities; they are constructed and recognized on the basis of reflection. Without possibilities, a person has no future and is already dead. This also applies to business: a company with no further possibilities is not a real company any more.

We use logic to find various possibilities for moving from a current set of facts to a possible new set of facts. If one cannot recognize possibilities, then one cannot act or plan for the future. Possibilities are logical constructs, but facts, value and language put limitations on the extent to which the logical possibilities can become real to people. Some possibilities are logical possibilities only and some are real possibilities – i.e. they can become factual.

Without logical tools, thinking would reduce to association, and argumentation would reduce to subjective expressions of preference. Such forms of associative “reflection” do not establish a horizon of possibilities, as distinct from impossibilities. Thus, in contrast to social constructivism, we consider the logical or systematic construction of models as important for organizations to operate and coordinate their activities efficiently and effectively. The world is not limited to a string of facts – that applies to the past only – but, with respect to the future, it is an open horizon of yet undetermined possibilities. Without logic it is not possible to recognize or to create future possibilities, to develop new tools or to discuss the idea of fairness, and additionally it is not possible to develop accounting tools.

Values

Humans have values, which are important in how humans experience and connect to life. Values provide objectives and thereby motivate people – i.e. they underlie their subjective preferences, feelings and likings. Thus, human values are subjective. We are not born with them, but they develop differently from human to human in a historical social process. It is values that make people act. However, values are not enough for human beings to

act; they also need possibilities – i.e. human beings will act when values and possibilities are integrated. One may have different possibilities for action, but only when one's values are integrated into at least one of the possibilities will one act.

A number of values are objectivized in the institutionalized system, such as modern values of wealth, influence, recognition and fame or traditional values such as character, benevolence, courage and honesty. Nevertheless, these values are instrumental only. True values, values that are not instrumental for some other purpose but for the purpose in themselves, are subjective. They are personal. They make life good for people, and the only standard for this is the life experience of people. What that is depends solely on what people like, love and find lovable, not as a means to something else but just because it is as it is.

Communication

Without communication, only individual reality exists: there is no intersubjective socially organized reality, and neither companies nor institutions can be managed. Such disconnected “reality” is, however, not possible, because communication is a necessary aspect of individual reality. Integration takes place in a social setting and communication is establishing the intersubjective reality. Communication, including the fundamental tool of language, enables people to cooperate and management to access the subjective worlds of the values and reasoning of employees. Communication allows the objectivization of values and reasoning. It therefore establishes social values and implements a social reasoning that, for example, determines which management control concepts and models may be used in the company. Thus, through human interactions, a personal objectification takes place whereby values and institutional rules are determined. For individuals, as well as for organizations, it is important that their respective values are integrated with the available facts and in the logic.

The body of accepted perspectives, arguments and concerns, which is used to control the common reason or decision making of an actor's communication, is captured from rhetoric by the notion of “topoi”. The subjective topos is the communicative expression of the mode of thinking by which an individual perceives, understands and interprets reality. The interactions of an organizational unit of human beings and their joint actions are controlled by a social topos constructed from the joint ideas and concepts,

interpretations and patterns of thought of a group of human beings. Social topoi govern the construction of organizations and institutions.

Integration

Full integration constitutes the reality of the actor in the world. It implies that the actor belongs to the world and that the world is the world of the actor. To the individual the world appears as reality, due to the integration that opens the time horizon and enables the individual to pursue his or her values. The actor constitutes him/herself as an actor by managing his/her relation to the world in a way that ensures full integration. Without a proper integration, there is no actor, neither individual nor organizational. To support the construction of functioning actors, society develops a set of institutions, media and culture.

The world in itself – i.e. the world without actors – is, from an epistemological point of view, an abstraction. It is a world in which the operation of natural laws by themselves determines what happens. This is the physical world. The category of possibility is not relevant there. Possibilities are relevant for actors, to enable the actor to pursue his/her values according to his/her own intentions. The construction of possibilities sets the actor free to pursue his/her values.

Summing up, we assume that social reality includes four integrated dimensions: facts; logic; human values; and language. These are all important in constructing social reality (Nørreklit, Nørreklit and Israelsen, 2006). If one of them is omitted the relation to the world will not function, and the actor will be unable to function. Therefore, the reductivism inherent in one-, two- or three-dimensional paradigms means that they are inherently problematic as instruments for social decision making and control. The simple-minded one-dimensional type of thinking that attempts to identify the one cause that is able to solve all problems is notoriously problematic. The traditional paradigm of realism has a severe reduction in its concept of reality, in that it has no proper place for the subjective elements of meaning and value. Consequently, it cannot sufficiently solve the problem of validity; indeed, it will distort validity. By integrating facts, logic, values and language, realism, rationality, hermeneutic (interpretive) and language philosophy can also be integrated. In practice, these dimensions are often more or less integrated; in theory, however, they are missing. The implication of this paradigm for the epistemology of accounting performance measurement will be explained further below.

Pragmatic constructivism versus mechanical realism

The paradigm implicit in traditional accounting performance measurement frameworks can be characterized as mechanical realism. It considers reality as a set of independent facts. These facts have to be recognized and measured in the description of activities and performances. The relations dominating this world are of a causal nature. The task of performance measurement is to identify and measure the causative factors. Statistics therefore become dominant. This concept of causality is one-dimensional. It looks for the one decisive cause, and if that cannot be found it looks for a combination of causes. This means that, as, for example, with the balanced scorecard (Kaplan and Norton, 1996), the basic relationship between activity and action and its result, effect or performance is considered as a question of cause and effect.

The problem with mechanical realism, however, is that this approach creates validity problems. The reason is the inadequate concept of reality upon which it is based. This inadequacy leads to an improper construction of the concepts of fact and causation. The concept of reality is not a concept about a world independent of human beings, but a concept about the actors' realistic, functioning relationship to the world. This shifts the whole perspective from mechanical realism to pragmatic constructivism, namely to the construction of the functioning relations so that individuals and organizations can constitute themselves as actors. The integration of fact, logic, value and communication creates the performance of the activity, but not in the way that a moving billiard ball causes another billiard ball to move or a pilot causes a plane to fly by pressing the relevant buttons. Reality is not the construct of one-cause. The causal ideal of one-dimensional leadership cannot be used to control events. Accounting performance measurement frameworks, based on statistically determined cause–effect relationships, are unlikely to capture the complexity of the behaviour that underlies performance. The complexity of integration involves not only measurements but also judgements concerning the adequacy of measurements and the sustainability of integration.

Paradigms, as fundamental beliefs about the world in which we operate, imply, either implicitly or explicitly, that a particular concept of truth applies to that world. From different paradigmatic bases it is, therefore, possible to conceive of truth in different ways. However, to operate with an internal consistency and sound external justification, a profession such as accountancy should adhere to a truth concept that is compatible with the

paradigm it adopts to explain, conduct and promote itself. The accounting profession has a dilemma in this regard. Its traditional paradigm of realism assumes a tangible, real-world, truth substantiation for the performance information that it produces. Such substantiation cannot be found for many components of accounting measurements. Consequently, the profession opens itself to criticism from those who identify these disparities. To assume a defensible position in society, accountants need to base their discipline on a concept of truth consistent with their prevailing paradigm. The following section explains how, using the paradigm of pragmatic constructivism as a starting point, this might be achieved.

Truth and valid measurement frameworks

There is not one question about truth. There are several questions that a theory of truth should deal with. What is truth? What are the criteria of truth? What are the conditions of truth? These questions are addressed differently from various philosophical perspectives. The framework of pragmatic constructivism sets the following aspects as involved in truth: the pragmatics of the actor perspective; the surrounding world; and the integration of the four dimensions in the relation between actor and world. This section discusses the theoretical conditions for valid truth in performance measurement. By integrating ideas of truth we develop an integrative theory of truth that enables us to point out theoretically the problems of validity. The result is a learning theory of truth in which the learning circle – i.e. the interplay between pragmatic truth and proactive truth – forms the basis of the learning process.

The pragmatic perspective on truth

According to pragmatic constructivism, the four dimensions are integrated with respect to a pragmatic perspective – i.e. to enable the existence of actors in the world. This must be reflected in a valid concept of truth. According to a pragmatic theory of truth, a statement is true if it leads to successful actions – i.e. if its future predictions hold true. This presupposes, of course, that a statement has such future implications. This presupposition fits the paradigm of pragmatic constructivism at the point in the construction at which all our activities participate in the construction of a valuable future.

A drawback of the pragmatic concept of truth is that one can only know the truth after having tried to apply it in action. One cannot know the future before it has become the past. We would, however, like to know a truth in advance, before we take action. Furthermore, this implies that there is a paradox in the traditional pragmatic definition of truth: it leads to an infinite regress. If we say that the truth of a statement is always a truth about what happens in the future, then we will never know the truth. Suppose a statement lets us expect that something, X, will happen. We cannot test whether it happens, because the truth about whether X did happen or not is simply a further truth to be tested in the future.

The phenomenological grounding of truth

Establishing a factual ground is essential to pragmatic constructivism. To do so, careful observation of the phenomena is necessary. The world around us, however, is infinitely complex. No matter what part of the world we study, there is no end to its complexity of phenomena related in all kinds of ways, and loaded with all kinds of meaning, values and possibilities to the observing actor. Moreover, as if this is not enough, all the phenomena and all their relations, meanings, values and possibilities change in many directions simultaneously. This all implies that there are endless values and possibilities for the actor, but it also implies that it is difficult to know with any degree of certainty what the essential relations are that the actor can actually rely and therefore focus on. He needs to consider it all carefully. However, what we can express in our language and what is documented in our data and summed up in reports is always highly simplified. Our language has very few concepts by which we try to express what we consider to be essential in the infinite complexity of the world. Thus, there is always a question of whether we link our limited language to the infinite diversity of the world such that it reveals the dynamics of the world in a way that is most relevant and sufficiently acceptable to the actor.

From a phenomenological perspective, truth is concerned with uncovering the complexity of the world, not as it appears dressed in language or theoretical concepts but as it is. The essential aspect of truth is that everything is in the open and nothing is hidden from the observer. This is a condition for an adequate construction of the relation between world and language. Therefore, truth is the *uncovering* (alethaia) of the phenomena that make up the world we are studying (Heidegger, 1927). The unclouded appearance of the phenomenon gives us the truth. An observer cannot base

his/her opinion on some representation of the “things” but must confront “the things themselves” (Husserl, 1913). Thus, phenomenologically speaking, truth is something apparent in the direct contact with these things or phenomena. The basic error according to this perspective is not to study the real world carefully with an open mind but to presuppose that given pre-definitions and pre-understandings are true and adequate. We must, so to speak, put our pre-understanding in brackets (epoche) (Husserl, 1913) or challenge our pre-understanding (Gadamer, 1960) if we are to experience the real phenomena of the area we want to study.

We agree that this acquaintance with the real world is a condition for real truth. Without acquaintance with the phenomena of the real world our statements about it become invalid. However, it is a condition only. In this complexity of phenomena we must determine pragmatically significant relations that can be used in a construction of reality, and we must put this knowledge in a conceptual intersubjective form – i.e. we must create the language, the concepts and the models necessary to formulate true and reliable statements. In the following we describe the elements involved in creating linguistic and theoretical expressions of the truth, which we call “proactive truth”. The guiding criteria will be correspondence, coherence, interpretation and acceptance.

Correspondence and facts

In order to establish something as a fact it must be dressed in language. The correspondence theory of truth considers truth to be a form of match between a statement or an assumption and the relevant part of the world. That a statement is true means that the things in the world are as they are expressed in the statement. If things are different from how they are described by the statement, then it is a false statement.

The correspondence theory of truth is concerned with establishing a relation of matching between statements and phenomena in the infinitely complex world. This is typical for the common-sense positivism/realism lying behind mainstream theories. Concepts are supposed to refer to specific types of phenomena. A model is a structured set of concepts that is used to create statements and theories to express a situation or set of relations between the phenomena. Truth becomes an issue first when the concepts are used to express statements and create theories about the relations within the subject matter. Truth is considered to be correspondence in the sense that the

true statement somehow mirrors the actual state of affairs (Wittgenstein, 1922; Mautz and Sharaf, 1961).

We need, of course, to prove this correspondence or match in a pragmatic way. A statement matches the world if one can accept and trust it without becoming disappointed when acting according to it. The expectations it gives us will be fulfilled if it is true. The match is true if its pragmatic consequences hold true – i.e. if the expectations it gives us are fulfilled. Nevertheless, we say it corresponds now and is true now although the expectations that prove it relate to the future.

The epistemic platform

The correspondence idea about truth theoretically lifts truth out of the subjective sphere of acquaintance and dismisses any opinion about a statement as irrelevant to its truth. In practice, however, this is not so simple. Statements about the world seem to have a subjective basis of evidence even though the truth itself is independent of such subjective relations. We must, therefore, find ways to reduce and overcome the subjectivity in our knowledge of the truth. This we do by creating criteria for applying our concepts and systematically organizing this criterion-based information in an epistemic platform. This provides us with basic intersubjective information, out of which we construct what we consider objective truth. The epistemic platform is an intermediary between the real phenomena and our theories, statements, concepts and models.

Although phenomenology and positivism are both concerned with the empirical world, the positivist correspondence idea of truth is concerned with the use of language, while phenomenology is concerned with awareness of the realm of phenomena. In positivism, truth is not about confronting “things themselves” but about collecting empirical evidence or data that document the true statements. Direct observation of phenomena is subjective and not always reliable. They are vague, they can be distorted by memory and they are imprecise. To overcome these problems, an epistemic platform of evidence and data is created as a medium between the realm of the phenomena and the level of concepts, models, theories and statements. Even grounded theory, which proclaims to be a phenomenological method, constructs a phenomenological platform in the form of interview data and uses it to construct concepts, models and theories. The concepts are not constructed by a direct confrontation with the phenomena but, indirectly,

by analysing the data collected. Although it is complicated and even expensive to create such a platform, it has many advantages:

- the evidence or data collected is intersubjective and enduring;
- the evidence bears some kind of reliable witness about the phenomena because it is somehow produced by the phenomena themselves; and
- the evidence and data can be collected and analysed systematically.

Despite these advantages, the use of an epistemic platform also has some drawbacks. It creates a wall between the observer and the phenomena. This is a major basis for validity problems, because the direct concept–world relation, which so much of everyday life is based on, is replaced by two indirect links: the link between concept and data and the link between data and phenomena. The attention of an observer or controller may be limited to the relation between the data and their conceptualization, which easily creates a situation in which a controller is out of touch with reality.

The direct relation between concept and practices is replaced by a double relation: the relation of the evidence or data to the observed phenomena and the relation between the concepts and the data. It is the combination of these two that now constructs the reference to reality. Validity problems occur if there is (1) insufficient grounding – i.e. there is an unsatisfactory phenomenological basis for the collection of evidence – and (2) misleading conceptualization – i.e. there is an unsatisfactory analysis of the concepts in the theory. If the phenomenological grounding is insufficient, then the evidence and data are likely to be misleading as expressions of what is going on in the area under investigation. In statistics, this means that the data are not representative. In grounded theory, this means that the interview material does not adequately cover the field of study. If the conceptualization of the concepts used to create the models and theories is unsatisfactory, then the link between the data or evidence and the statements or theory is misleading due to the misleading operationalization of the concepts. In grounded theory terms, we might say that the procedure to create concepts out of interview data by coding has not been performed adequately.

Some phenomena, such as physical phenomena, are ontologically objective; in contrast, some, such as personal perceptions, feelings, thoughts and values, are ontologically subjective. Correspondence and the ability to obtain objective knowledge is important not only when we talk about ontologically objective phenomena but also when ontologically subjective phenomena are concerned. Although it is easier and more common to use the epistemic platform to create information about objective phenomena because the

subjective phenomena are often very sensitive, both are important to the understanding and analysis of performance systems. Thus, we often combine direct observation of phenomena with the indirect reporting of objective phenomena.

One of the consequences of this transformation of the infinitely complex world into simple language is that the linguistic truth is an abstraction that simplifies the world. Our language is very limited compared to the infinite complexity of the real world. We are forced to express the truth about the world by means of just a handful of concepts. We even attempt to use this truth to enable us to measure and control events.

The epistemic platform is used for two complementary purposes: it is used exploratively to create a basis for conceptualization and it is used to create a basis for ongoing observation of conceptualized phenomena. We consider the conceptualization process to be a process of abstraction, very much in accordance with grounded theory. We develop a set of saturated concepts that create a model of the phenomena. Based on the model, we further develop specific hypotheses and theoretical statements about essential relations.

At this level we may have sufficient knowledge to create measurement instruments that can be used to create precise, relevant information. Measurement is a special way to create data for the epistemic platform. When observing the infinite complexity of the world we need structure; when we start to know how things interrelate, however, then we soon need precision in order to determine performance precisely and create increasingly sophisticated results. Measurement is based on established standards that are presumed to express the concepts/variables measured in numerical terms. The overall purpose of using measurement is to obtain a higher degree of precision. Therefore, measurement is especially related to the factual dimension. We cannot measure the values and subjective world, although we can create indicators of it. We cannot measure meaning in communication and we cannot measure possibilities. This means that the meaning of measurement is always a question of interpretation. A measurement itself is true and correct if it is performed according to adequate procedures, but, when we use it as an expression of social reality, validity depends on how the other three dimensions are handled. We call this handling of the other dimensions the “dimensionalization” of measurement. Measurement without proper dimensionalization lacks meaning. In this form it is a means that can be used by people in charge to any purpose, to reward or to blame. Dimensionalization means integrating possibilities, meaning/communication and values.

The logic dimension and theoretical coherence

The coherence theory of truth is mostly associated with rationalism (Descartes, 1988 [1628/9]; Kant, 1999 [1787]) due to its focus on theoretical relations, definitions and analyses of inference and consistency. The criteria of coherence and consistency are not, however, restricted to a rationalistic perspective. For instance, grounded theory allows the analysis of theoretical coherence in a phenomenological context. Concepts such as coherence and consistency are mostly considered as characterizing theories and other symbolic systems. However, not only do we want to have a system that is consistent, we also want the system to reflect and analyse the phenomena with respect to their pragmatic coherence. Pragmatic coherence means that the phenomena are sufficiently integrated to allow the actor to function and realize his/her values and goals.

Social reality is itself a reflecting system because of its integration of the logic/possibility dimension. A system of activities is a social system. It is coherent if it maintains its ability to perform – i.e. if it is able to continue realizing the purpose it serves. This means that it is incoherent if it implies the destruction of some of its own conditions, or if it creates results that are irrelevant for its purpose. The question of pragmatic coherency is a question of organizing a coherent flow of activities relative to the demand, need and desire for them.

The question of coherence comes from the fact that things have consequences in theory as well as in practice. In practice, one phenomenon follows from other phenomena. In theory, this is expressed as one type of phenomenon following from the combination – or, rather, integration – of other phenomena. If things are incoherent then the consequences are chaotic and unpredictable, because the various consequences conflict and thus destroy each other. This is the case in incoherent practice as well as in incoherent theories.

A system must be formally consistent – i.e. it must not imply contradictory statements. Consistency is not enough, however. The task of the conceptual system is to enable us to analyse phenomena with respect to enabling the actor to function – i.e. the system must enable us to analyse the coherency in the real flow of activities with regard to the creation and maintenance of a functioning integration. In this way a pragmatic coherence is created. To do this, the model must identify the conditions of harmonious flow in order to identify possible upcoming coherency problems. As knowledge of the pragmatic flow increases, we integrate this knowledge into the conceptual

structure, thereby gaining improved analysis of the possibilities of the situation.

The very point of analysing coherency is to obtain detailed knowledge concerning our possibilities, how to realize them and the consequences they have for our values. The point in discovering stable relations is that it enables us to organize our activities and behave systematically to fulfil our values. It is not the task of our analysis to come up with the one and only possible procedure, nor is it intended to describe the world as fully determined. It is to enable us to analyse what we can do, and it is for our values to determine which of the possibilities we want to chose. An analysis that claims that there is only one possible way may be true, but our world is not normally this simplistic. People should not be deprived of the ability to choose.

Acceptance and value coherence

While plenty of methods have been developed to analyse correspondence, coherence and meaning, it is difficult to find convincing concepts and methods that are appropriate for uncovering and analysing values. We need first to point out the difference between what we may call “true values” and “instrumental values”. The true values serve no purpose. Instrumental values are valuable only insofar as they serve the realization of true values. The social construction of values such as power, wealth and fame are instrumental in motivating people to maintain the order of society. The very purpose of society itself, however, is to help people organize a valued life. The structure of society is not a purpose or value in itself: its purpose is to serve its people. True non-instrumental values are subjective. They concern the love and feelings of people and create the basis for the motivation of people’s activities. We do not know what to expect from people if we have no understanding of their feelings.

However, the subjectivity of true values makes it difficult to observe them objectively and report them in the epistemic platform. We consider subjective values to be linked to the emotions and feelings of the individual. Consequently, uncovering values involves not only language but also a perception of the person’s feelings and emotions. Through dialogues and discourse analysis, one may obtain knowledge about what a person’s values are – what he/she likes and dislikes, what motivates him/her, what he/she considers to be ethical. However, the person may not tell the truth about his/her values or he/she may not know his/her own values. Growing up in a social context, one learns and adapts to the social norms and values of the environment. Although discourse analysis may uncover conflicting discourses and thereby

suggest conflicting values, language cannot stand alone. We suggest that values are located within the actor, and what is expressed in language must point to something within the person in order to be true.

Since the basic meaning of things concern their values, this is one of the places where misunderstanding easily occurs. There is only one way to overcome this problem, and that is by creating an open environment in which one can have confidence that people are expressing their values and feelings. In such an environment, one can use a signal of acceptance from an individual as an indicator, not only of his/her knowledge and skill but also of his/her values. One may also use his/her hesitation to accept something as an indicator that his/her values and judgement disagree.

Habermas (1990) refers, in his communicative ethics, to the force of the rational argument, and is concerned with conditions in which this can outweigh an unequal distribution of power in discourse. We have a similar ideal. However, the essential problem is not the rationality of the argument as such but the openness and respect for the subjective acceptance or rejection signals – i.e. the subjective value signals of the individual participants. Such a situation presupposes sufficiently strong social norms to make people in power refrain from misusing this power to create fear, and cause people with less power to hide their true values. If such norms do not exist then people react with fear to the people in power, thus jeopardizing the validity of their value statements.

Communication and social coherence

In a social construction there are always a multiplicity of actors, and therefore also a multiplicity of different conceptualizations of the subject matter. Each of these must be theoretically coherent. The flow of activities has its own coherence, which is more or less matched by the conceptualizations of the involved actors. This structural match between the various conceptualizations forms the overall coherence, the social coherence of the operating construct. This match is an expression of how well the actors communicate and understand each other – i.e. it is a match between the subjective topoi and the social topos. External or social coherence is a condition of the clear meaning of truth. If people understand the concepts differently, then there is no consistency in the truth claims. It is common for people to interpret the same words and concepts rather differently, especially in interdisciplinary contexts such as the operation of a company. A condition of validity in truth assumptions is, of course, that the meaning is correctly

understood. The intersubjective character of the truth claim must be established. To this purpose, we have a broad range of tools to establish understanding and achieve mutual comprehension. Reflecting the creative nature of actors as individuals and as components of companies and organizations, as well as the complexity and changing nature of the realm of phenomena, such understanding cannot be achieved through rules and predefined concepts only. Creativity, complexity and change are also concerns for the meaning of communication. Necessary changes are made to the meaning of concepts in the integration process. Sameness of meaning cannot be presupposed but must be established. Therefore, it is necessary to use dialogue and other forms of mutual open communication to establish mutual comprehension in a creative performing environment.

Thus, we distinguish between internal theoretical coherence and external social coherence. Internal coherence is a condition of the operating conceptual system of an actor, while external coherence is a condition of the relation of an operating conceptual system to another operating system. Since actors are either individuals or organizations of individuals, we also have conceptual systems at different levels. This implies that the internal/ external distinction is relative, in that the internal coherence of an organizational actor reflects the external coherence between individual actors in organizations. To organize this complexity and utilize the many forms of knowledge, and the skills of the individual actors, the communication and interpretation processes in companies must regulate and coordinate the interplay between different types of conceptual perspectives. This process is supported by the functional structure of the firm. While theories related to functional perspectives enhance the internal coherence of specific activities, they leave the external coherency to an undefined interdisciplinary perspective. To overcome this problem, various more holistic theories attempt to create a basis for creating external coherence.

The overall ideology of society further creates institutions to regulate the way objectivizations can be created and function. This process of objectivization and institutionalization creates relatively stable conditions within which actors can cooperate. Institutions and objectivizations are themselves conceptual constructs, and thus subject to the criterion of internal coherency. However, as institutions they set the conditions with which all operating activities must be externally coherent. The institutions are, therefore, supported by the legal system. For the firm to be legitimate it must cohere externally with the institutional conditions, very much as the individual actor in a company should cohere externally with the objectivizations of the firm. In order for a truth about the situation of a performing actor to be valid, it must

be externally coherent with the institutional conditions. Internal coherency is not sufficient. The legitimization of the firm is based on the availability of sufficient relevant data in the epistemic platform.

Social constructivism, which assumes that human thoughts, discourse, agreements and concepts create the world, is a one-dimensional approach, focusing on communication only (Nørreklit, Nørreklit and Israelsen, 2006). The mistake of social constructivism is not in assuming that facts are socially constructed. Unknown and undefined facts are not facts. The mistake is to conclude from this that that which is a fact is socially constructed. For instance, that something is a chair is a socially constructed fact, but this does not mean that the chair is socially constructed by the use of language.

From the pragmatic constructivist perspective, we destroy the actor if he/she is controlled by language usage and meaning that is not also the meaning attributed by the actor. The establishment of mutual understanding through a so-called “hermeneutic” circle (Gadamer, 1960) seems much more important to the understanding and control of social processes than formal hierarchies of control and command, especially if no dialogue and mutuality exist in the vertical structure of the hierarchy.

Proactive truth and the learning approach to validity and truth

The different theories of truth mentioned above inform us about different aspects of truth. However, none of them can stand by itself. We need an integrated theory of truth. Correspondence concerns the factual aspect while the various forms of coherence inform us about the other aspects – the logical, communicative and value aspects; coherence is the logical aspect. Interpretation (hermeneutics, discourse analysis) is the primary methodological approach to clarify how language, communication and conceptual systems comprise the integration of factual, logical and subjective aspects of truth. The phenomenological theory concerns the quality of our acquaintance with the real world, including its objective as well as subjective aspects, by challenging our prejudgement and uncovering hidden phenomena.

We now present the pragmatic constructivist theory of truth in terms of a distinction between proactive truth and realized or pragmatic truth. Proactive truth is truth we achieve when we combine the criteria of correspondence and coherence. Pragmatic truth is the realized result. While proactive truth gives us expectations and anticipations, pragmatic truth tells us whether our expectations were realistic. The proactive truth is based on

our knowledge, concepts and calculations. It relies on the information available in our epistemic platform. Pragmatic truth is an expression of our experience as actors and observers in the phenomenal world. For instance, we may estimate the value of a commodity, say a car, by trying to identify its market value. This gives us the proactive truth. It is not the real or pragmatic truth. We suggest that the real truth depends on the real condition of the car as well as the market, and these are not known precisely. The real truth also depends on the ability and conditions of selling. We find the pragmatic truth only by actually selling the car. For pragmatic constructivism, the proactive truth is based on an abstract system that relies on an epistemic platform. This system must continuously be tested and grounded by confrontation with realized pragmatic truth in order to maintain its validity and reliability.

The difference between proactive and pragmatic truth underpins the fact that we learn and continuously improve our knowledge system. If there is no such learning process in operation, then we cannot obtain reliable price information. If I sell the car at a lower price than the market value I might, for instance, learn that my ability as a salesperson is below the market level for this product.

How integration works cannot be demonstrated a priori. There is no absolute truth about the integrative construction of reality. Each culture, each history, unfolds its own specific form of integration. The integrated construct is a pragmatic enterprise in which we establish ourselves as actors. The integration depends on the prevailing social conditions, cultural and institutional, which must be learned in order to become a recognized actor. The learning takes place through the pragmatics of truth – i.e. by opening the perspective of time and comparing proactive truth claims with the pragmatic truth. If activities do lead to intended results, then the interpretations used are considered true. If the results deviate from expectations, then we need to improve our understanding of reality – i.e. of the operating integration. We are always able to improve our understanding of the character of operating forces. The phenomenological grounding reveals the phenomena as they appear, loaded with dynamics, meaning and perspectives. The structured reality, the pragmatic coherence, what are important forces and what not, emerge in the ongoing learning process between the proactive and the realized truth. The measurement points selected have no guaranteed stable meaning. The ground is not stable. The physical facts are stable for a period, but their meanings and significance are not. Consequently, proactive truth cannot be reduced to a mechanical registration and calculation. Proactive truth also demands involvement and judgements.

Designing a valid performance measurement system

Below we suggest some steps to be followed in order to establish valid performance measurement from the perspective of the *integrated learning theory of truth*. The steps in this endeavour are as follows.

- (1) Understanding the task of a performance measurement system.
- (2) Grounding – the performance system as a phenomenon.
- (3) Modelling and dimensionalizing the epistemic platform, bearing in mind:
 - (a) internal coherence;
 - (b) external coherence: company and institutional coherence; and
 - (c) the epistemic platform: truth versus beliefs.
- (4) Proactive truth, real truth and pragmatic learning.

Although, for convenience, these are presented in an analytical framework, the steps are not separate and linear steps but mutual and interactively linked together. We consider the steps to be generally applicable for different types of performance measurement systems solving various purposes. However, the steps have to be considered properly for each purpose. In our description below our main focus is on accounting measurement systems for the purpose of performance identification, and accountability assessment both within and outside the firm.

Understanding the task of a performance measurement system

For a firm to survive it must continue to integrate facts, logic, values and possibilities successfully – i.e. it must create and ground possibilities and values in the facts of the social organization. If disintegration takes place, then activities become incoherent and the system dies. Therefore, maintaining pragmatic coherence in its activities is the way to ensure continuing integration. Given the aim of obtaining certain results, pragmatic coherence focuses on whether the relevant activities and actions match or complement each other in such a way that the company can realize its purpose. The systemic task of performance measurement is establishing knowledge about and control of the degree of pragmatic coherence in the flow and performance of activities.

Grounding

As mentioned above, a performance measurement system (i.e. the units and concepts to be measured, the procedures of measurement and so on) is a

reduction of the performance system and its situation as a phenomenon. The basic requirement for such a simplification to be of value in a measurement system is that it should be grounded in the reality and life world of the performing unit. The measurement and control system should not, therefore, be an a priori construct imposed on the organization.

The observational ground is the point of departure for a performance measurement system. Thus, the *performance-creating firm as a phenomenon*, as an experienced life world, lays the ground for a valid performance measurement system. The first step to understand this reality is to observe and study the firm and the purposes of its performance measurement system carefully and with great sensitivity, to uncover what is going on. Designers must identify what the business of the firm is – i.e. what the relevant physical, institutional and organizational details are and what the dynamics of these phenomena are – so that facts and possibilities can be established. Furthermore, the designer must uncover the purposes of a performance measurement system – i.e. do we need a performance measurement system for decision making, accountability, control, learning, etc., and how are these purposes perceived? Uncovering the business of and the need for performance measurement also reveals the understanding and misunderstanding of the participants, suppressed observations, integrated and non-integrated motivations and values, conflicts between groups, and the level of knowledge and skill. As a phenomenon, the company is, of course, infinitely complex. What interests us is the functioning of the company in relation to its environment. The observation of the knowledge of facts, of possibilities and of values and their integration with communication in this life world are essential for understanding the phenomena as an expression of reality.

The observer must go directly to the world he/she wants to know the truth about and uncover things that are hidden. Overall, the methodology for gathering information about the firm as a phenomenon is crucial for grounding the model in physical and social facts, in logical possibilities and in human language and values. We suggest an interactive and dialogue-based method as an important tool in this process.

Modelling and dimensionalizing a performance measurement model

The designer collects his/her information in a preliminary epistemic platform consisting of observational notes, interview summaries and other relevant material. Based on this material about the performing unit as a

phenomenon, he/she constructs a *conceptual model of the performing unit* by an iterative process of abstraction, whereby he/she asks theoretical questions about the emerging concepts in order to create a model of saturated and interrelated concepts. If his/her material does not provide answers, he/she will have to go back and make additional observations, ask people additional questions or obtain relevant material. The purpose is to create a theory that reflects the performance system in such a way that it can be used to observe and improve its coherence. To do so, all the aspects of pragmatic coherence must be reflected. The various forms of coherence reflected below are aspects of the pragmatic coherence of the performing system.

In this way, the phenomena are conceptualized and put into language. Instead of trying arbitrarily to implement a theoretical model, a model of the assumed reality of the firm is created, based on abstraction from the phenomenological material, much in the manner of grounded theory (Strauss and Corbin, 1990), which may be combined with the actors' perspective (Arbno and Bjerke, 1997). The performing unit as a conceptual model is a basis for structured communication about the nature of the system. A model that is not properly grounded in the performing unit as a phenomenon tends to overlook phenomena that influence the working of the system.

Internal coherence: coherence of concepts and theory

Internal coherence is a property of a model or theoretical system. It has two aspects: theoretical coherence and operational coherence. In order to create a measurement system, we observe a need to have a system of general, abstract concepts covering the phenomena to be observed and measured. The validity of this system of concepts depends on the relation between two complementary ways of determining the meaning of the concepts (see figure 9.1).

On the one hand, we define a concept vertically by formulating criteria for observing and measuring the phenomena it covers – i.e. by establishing rules and procedures for observation and measurement. Measurement variables are part of the vertical or operational definition of the concept. Measurement information is a substantial part of the epistemic platform of performance measurement. Often a whole theory is constructed around a measurement variable in order to understand a variable and make the measurement instrument function reliably. Normally it takes time and experimentation

Theoretical level	Concepts, model, theory, theoretical relations
Epistemic platform	<i>Data, measurements, procedures, criteria</i>
Phenomenological ground	<i>The phenomena, the performance system</i>
Horizontal (theoretical) definition: concepts are defined by other concepts Vertical (operational) definition: concepts are defined by reference to phenomena	

Figure 9.1: Epistemic levels

before reliable measurement procedures have been constructed and the platform has become solid. In order to make a concept operational, we must define measurement standards and rules, including the conception of a dimension, the definition of the units and the measurement scale, the expression of the units numerically and the description and application of an operation that discovers the number of units in a given object (Sterling, 1970). In some area of measurements we have problems establishing a concept that can also be made operational. For example, we may have a meaningful concept of such features as goodwill and social capital, but we are unable to make the concept operational.

On the other hand, we determine our concept horizontally by relating it to and defining it with other concepts, including the concepts of our conceptual model. The vertical definition described above is the basis of correspondence, and links the concepts, model and theory to the world. Measurement is part of this linking, aimed at providing this link with precision. However, in itself reference to reality means nothing; it is only a form of pointing. For measurement to have meaning it must be integrated in a theory – i.e. be related to concepts that are theoretically defined by means of other concepts. Only in the theoretical context of a model with specified relations can we produce statements in which values measured have meaning as expressions of performance measurements. The horizontal definitions specify the relations in the model. The relations must be made specific. If we do not specify the relations, then we can formulate no significant statements about the performance system but only state something general, such as that all the elements of the system are interconnected. Such a model generates no information that can be true or false. When these conditions are met, however, then the model has become an operational theory.

The statements specifying the relations must be internally coherent. Otherwise, a measurement may mean contradictory things. In measurement, a variable is given a specific – usually numeric – value. For this to have meaning, not only must the variable be part of the operationalization of a concept but this concept must be part of a conceptual model of the performance system. Measurement that is not part of such a model is arbitrary. If it is arbitrary, its function mainly becomes a political instrument in a power struggle that is irrelevant to real performance measurement, and its effect on performance may be negative. Thus, when measurements are used, it is essential to ask for the theory or model that explains the point of the measurement. If there is no such model, then measurement cannot be considered to be anything but a set of intuitive indicators with very subjective value only. The point of measurement – an increase in precision – will be lost.

There is a risk that one merely picks up and finds something that is measurable. Such a procedure results in vague and meaningless measurements. We can attach a number to anything; but that does not mean that we measure anything we give a number, or that anything with a number attached is relevant. A calculation model alone does not make a figure meaningful. There are many models that were constructed without a proper operational theory. For example, intellectual capital includes a lot of measurements, although it is often unclear what the measurements say. Similarly, the balanced scorecard is a loose conceptual framework, which makes the concepts open to interpretation and the linkages between the concepts ambiguous and thus not suitable for valid measurement.

The theoretical coherence of the model relates to the problem of operational coherence. This means that the measurements must measure variables as they are theoretically defined by the model to be – i.e. the vertical and the horizontal definitions of a concept must coincide. Otherwise, our measurements will relate to something different from what we theoretically say we measure. To avoid this is precisely the classical problem of validity. Consequently, analysing the validity of a measurement system includes a validation of the concepts in relation to the phenomenon as well as the theory embedded in the framework defining the measurements.

When an internally coherent theory is established, then we are able to analyse what is theoretically possible relative to the information given by the measurements. Measurement should establish the theory with facts that allow the theoretical possibilities to be identified. We have a language with which we can communicate about the unit.

Our concepts involve subjective judgement, which may cause our interpretation of measures to change in space and time. Moreover, the environment of the measurement process includes social and political spheres of interest that influence the measurement situation. It is to minimize these problems that we develop specific criteria – i.e. specific language rules and procedures, which apply to the measurement of phenomena independently of who is considering them. Without precise rules for measurement, measurement can itself give no precise information. Often whole theories are constructed in order to make a specific form of measurement reliable and precise.

Although everything can be given a number, it seems that there are things that cannot be measured adequately, since measurement is itself always part of the system. Adorno, Dahrendorf and Pilot (1991) argue that culture can never be measured. Meaning is lost as information becomes so highly reduced. Since measurement is a powerful way to impose order and control on hazy thinking, it creates a temptation for controllers to focus on what is susceptible to measurement and ignore or reconfigure the complexity and difficulties of the matter. Thus, the measurement becomes invalid; it does not recount something real and relevant about what is under investigation. The apparent neutrality and objectivity of measurements and the methodological procedure of using the epistemic platform has had the result that measurements are used on many phenomena without a proper understanding, and thus without knowing what they actually measure.

The stringent procedures in using measurement in natural science are often missing. Thus, we do not adhere to the point of view that, because verbal reasoning seems too “slippery”, exact results, even if faulty, should be preferred (Porter, 1995, 52–3). Nor do we agree with Lord Kelvin in his famous statement that “[w]hen you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfying kind”. We don’t need to know anything about research in order to measure, for instance, the number of printed pages or calculate the number of people in research groups. *From our pragmatic perspective, we know something about a phenomenon if we know how to use it successfully in our activities, irrespective of whether we can measure it.* A shortage of ways to measure a phenomenon adequately does not imply that one should not talk about and take it into consideration. However, the subjective nature of some value-dominated areas does make it difficult to develop standards for measurement. For example, Likert-type scaling is

Table 9.1: Activities involved in achieving specific goals

Goals	Activities					
	Production	Logistics	Engineering	Purchasing	Marketing	Finance
Productivity	x	x	x			
Inventory turnover	x	x		x	x	
First-time quality	x	x	x	x		
Delivery performance	x	x		x	x	
Number of suppliers	x	x	x	x		x

Source: Adapted from Grady (1991, 49).

subjective. It is difficult to know what the measurements relate to. They may provide indicators for values and emotions, but they may also be misleading. The subjective attitude is present in every transaction and recognized by the interacting actors, and may be reported and taken into account in the decision-making process as a hunch as to the direction in which things will develop.

External coherence

Internal coherency may be a condition of one operating conceptual system only. External coherency is a condition of the relation of an operating conceptual system to other operating conceptual systems. External coherency means that the internally coherent conceptual framework, its inner logic, must match the topoi of the various organizational actors or groups of actors as well as the constitutive institutions of society. Otherwise, internal cooperation in the firm and its interaction in society function badly. This is the task of establishing company coherence and environmental or institutional coherence.

Company coherence

Looking at a firm as being composed of multiple actors and groups of actors (Nørreklit and Schoenfeld, 1996), the conceptual framework must be linked to the subjective and social topoi within the company. For example, table 9.1 shows that different groups are involved in obtaining various specific company goals.

Company coherence means that the conceptual framework of performance measurements needs to be rooted in the factual possibilities, values

and language of the employees and/or groups of employees. Using an interactive method and grounded theory creates the basis for such a rooting. The design process is organized in such a way that the actors participate as linking pins in the design of a performance measurement system (deHaas and Kleingeld, 1999). While the design process is initiated from the top, the interaction among the various individuals and/or groups involved takes place as a dialogue. This gives the actors a bottom-up opportunity to contribute to the design of the performance measurements. A dialogue can contribute to uncovering strategic and operational opportunities and ensure the understanding of the system. It can be used to uncover factual possibilities and values and to develop the rules of the language game in the company.

The dialogue is also important in order to formulate the performance measures linked to the resources and internal commitments of the various actors or groups. It makes it possible to ensure that the actors within individual activities or constituencies have internal commitment to reach the ends by the given means. Rooting it in the internal commitment may contribute to ensuring that measures are realizable (Argyris and Kaplan, 1994, 91). Internal commitment means that the employees are motivated to be active, responsible and creative problem solvers. External commitment, in contrast, means that the employees find the motivation for their actions in rules and in the demands made by others.

The overall aim of the dialogue is to increase either party's knowledge, awareness, acceptance and understanding (Nørreklit and Schoenfeld, 2000). It may contribute to bridging differences of perception and comprehension, so that goal congruence and coherency increase.

Institutional coherence

By testing the model for external coherence with the institutional conditions, we analyse the logical possibility for the firm to function in the institutional and competitive environment.

Law, money and accounting are *institutional phenomena*, and they are also social constructs. They give rules and constraints for physical, individual and social possibilities. The constitutive rules state that some specific physical fact counts as a certain social fact in a certain society or group of societies. For example, there are phenomena such as cheques and money that constitute themselves through their collectively recognized status. A cheque enables the amount specified on it be transferred from the payer to the payee.

There is no hidden objective mode of existence, independent of any receiver behind the cheque, that the cheque reports and thus corresponds to. If it is not a fake it guarantees its own truth because it constitutes it. It is of the nature of an instruction such that “X counts as Y” or “X counts as Y in context C” (Searle, 1995, 28). Simply because of its collectively recognized status it constitutes reality.

However, Searle’s formula misses the point in the construction of social facts. It is not that a special piece of paper counts as a dollar that is the important social fact. That something is a dollar means, in itself, nothing. The important social fact is that a dollar represents a certain value, a certain buying power. This is a much more complicated state than can be expressed through Searle’s formula. The social facts depend on the social institutions and their credibility and ability to guarantee the status and value of the social phenomena they create. Searle’s formula can be used only to distinguish between a genuine dollar and a fake. This is important, but it is only important under the presupposition that the dollar has a value that cannot be created by simple agreement and a decision to let it count as a dollar.

Accounting procedures are based on institutional laws and regulations and generally accepted practices. We live in a property rights society in which accounting profit is the institutional formula that constitutes the selection criteria for being in business. As a tool describing the property rights of an entity, it has a collectively recognized status (Searle, 1995) in our property-based society. Institutional phenomena are determined through time by reflections, negotiation and power. As an example, accounting standards and practices are negotiated, and as is likely in all forms of negotiation, power plays a role in the outcome. There are human values behind the construction of these phenomena, but some values are more influential than others. Institutional phenomena change over time simultaneously with changes in the integration of facts, logic, values and language.

Accounting is not simply a construct created in line with the cynical intentions of the ruling class to install a false consciousness that legitimizes and reproduces the social system (Tinker, Merino and Neimark, 1982). Rather, it is an institutional fact that the organization of our society is heavily based on an economic logic, implying that *private* sector organizations are evaluated on the basis of financial measurements. The court of law, institutional standards and ethical codes imply that social relations and order are not simply a question of power and domination but also a question of human rights. By logic, we aim to develop standards for fairness and sensibility in our laws and regulations, which make criticism of the social

structure underlying the ideology of our society possible (Arbib and Hesse, 1986). Of course, there are ways of constructing society other than modern property-rights-based capitalism, which would require different logical constructs to regulate the resulting society. For example, in the former Soviet Union, companies for some time were evaluated on the weight of their production.

As accounting profit is a constitutive part of the economic logic of our property-rights-based society, accounting plays an important role. Management control and decision models are institutional language rule systems (logical constructs) that can help create economic profitability in a company. The external coherence theory in management accounting implies that the economic logic of the management accounting model should cohere with the institutional accounting model – i.e. the two systems should build on the same sort of logic. In order to be valid, management accounting models must cohere with the social construction of economic reality – i.e. the logic embedded in the management accounting model has to follow the rules of economic reality. In science, a common way of evaluating the validity of a measurement system is to observe the predictive abilities of the theoretical model in practice – i.e. realism assumes a rather objective point of reference, implying that an experimental, pragmatic criterion of success can be used. However, as accounting models are logical models, there is no such external objective referendum – although an external coherence theory of truth is at stake.

The balanced scorecard proposed by Kaplan and Norton (1996) fails to meet the criteria of external coherence. It assumes a cause–effect relationship between the variables relating to learning and growth, internal processes, customers and financial results. By assuming cause and effect relationships between certain types of behaviour (as reflected in the selected measures) and long-term financial results, these models are built on a generic relationship. This is evidenced by the advocacy of statistical testing to determine the cause and effect relationships. However, statistics used in this way do not relate to an underlying truth, and, indeed, may obscure logical fallacies that devalue the worth of the models. For example, the derivation of profit simply from customer satisfaction and customer loyalty per se is neither a necessary, nor a likely, occurrence. Profit depends on the levels of the costs and revenues of loyal and satisfied customers, and it is determined by financial calculus – i.e. a logical relationship, not a causal one (Nørreklit, 2000). On this basis it can be concluded that models such as these are not coherent with economic reality.

Our models for planning and decision making have to point to the institutional formula of accounting in order to say something real about what it should express. The implication of this is that those companies not using accounting and control systems can exist only in the short term. Even for a very small and simply organized company, it is necessary to use some kind of accounting system and to track cash transactions and the use of resources. We use accounting models to measure and report on these activities.

This does not mean that political, ritual and legitimating motives do not play a role in a company's decisions and acts (Meyer and Rowan, 1977; Edelman, 1977). Indeed, management models can serve as representations and ceremonies, leaving the managers with enhanced *ethos* and engendering the approval and acceptance of the social environment towards the organization. However, if economic considerations are not included then corporate economic profits will occur only by chance. Profitability in a company cannot be achieved automatically or only through legitimating acts. It is something that management and employees must work to create, and to do this they have to make economic judgements.

For planning and decision making the accounting tools involve financial calculus and budgeting systems, although decision making may use historical accounting information with respect to costing in order to generate prognoses. This is coherent with the way companies are evaluated. The problems with financial calculus and budgeting for decision making is that they have to rely greatly on subjective beliefs and judgements about future cash flows. Indeed, judging future performance and cash flows is, in many cases, a question of high complexity involving many aspects. Consequently, there is a need not only for a valid business theory but also for a valid method to judge the estimation of future cash flow. This will be developed further in the section below on pragmatic and proactive truth.

The epistemic platform: truth versus beliefs

Correspondence is established by creating the data of the epistemic platform and using them to relate the concepts to the phenomena. The nature and quality of the data is, however, essential. Figure 9.2 gives an overview of the ontological and epistemological nature of the data of different financial and management accounting phenomena. It shows what our data correspond to and whether our data are expressions of knowledge or beliefs. As the accounting tradition includes some physical quantities and social facts, the

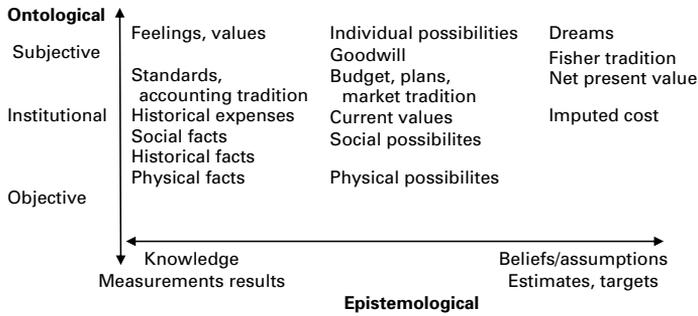


Figure 9.2: The ontological and epistemological nature of accounting data

accounting model provides, to a certain extent, knowledge. In order to create financial statements as social facts an elaborated set of rules have been developed to regulate the language of the financial statements.

In providing data for historical financial accounting performance measurements, we are concerned with the correspondence of the account information. The accounting report must be a true report of basic, relevant facts that disclose the essentials of the things going on – i.e. the physical and institutional phenomena. This is the realistic element: the fact that the reports and observations must point to something. The audit process reinforces the idea of a base of realism for financial statements, as it is a check that “the figures should accurately represent the physical phenomena that they are intended to represent” (Sterling, 1970, 14). The concept of “reporting a true and fair view” also indicates a form of correspondence. The statement is an expression about a financial unit, and it is true and fair if the situation of the unit is as expressed. The signature of the financial accountant is the public guarantee that this is the case. The financial situation of the company or the financial unit itself is, fundamentally, constituted by social facts and the financial statement. The financial statement is a social linguistic fact, which is supposed to express another social fact, namely the financial situation of the firm.

However, Searle’s formula for social facts – “X counts as Y in C” (Searle, 1995) – means here that, in a society C, the financial situation of company X counts as expressed in the financial statement Y, if Y is signed by a public certified accountant. This condition that “X counts as Y” is based on the credibility and the clarity of the rules by which the accounts are constructed, and on the trustworthiness of the revising work of the certified accountants. This depends on the ability of the physical fact, X, to maintain the special form that shows that it counts as Y. The trust further depends on the

trustworthiness of the legal system, that it will uphold the credibility of these rules. Only under these conditions can the financial institution create accounts as social facts.

Furthermore, historic cost accounting's limitations in terms of providing reliable information have long been recognized (e.g. MacNeal, 1970[1939]; May, 1972). Practical flexibilities concerning issues such as the depreciation of fixed assets, the valuation of stock, the establishment of provisions, the recognition of income, variation in the purchasing power of the measurement currency and the segregation of capital and revenue expenditure all make the aim of information objectivity difficult to attain (Sterling, 1979; Chambers, 1966). As Paton (1962) observed, "[T]he accountant is being constantly faced with the need for judgement. Accounting is full of estimate, of assumption. At times, unfortunately, the conclusions of the accountant must be well nigh conjectural . . ." Indeed, Sterling (1979) attributes these problems to the accountant's specific difficulties in relating information to the real world that it is meant to represent: "The primary problem of accounting is that our figures do not have empirical referents." Thomas (1969, 1974) reinforces this deficiency of conventional accounting practice through his comprehensive analysis of the "incurable" cost allocations that pervade financial accounting. While the establishment of regulations and standards can increase uniformity in accounting by developing rules for measurement and disclosure, it is unlikely that prescriptions of this type will eliminate flexibility. At best they restrict but do not avoid the problem of beliefs and, thereby, proactive truth.

In order to measure income and value more fully, alternative scholarly traditions rooted in neoclassical economics have evolved (see Fisher, 1906, 1930; Hicks, 1946; Edwards and Bell, 1973; Sterling, 1979). The Fisher tradition uses an economic concept of income: "[I]ncome is measured at the end of the period by adding up the discounted values of all net receipts which the managers then expect to earn on the firm's existing net assets and subtracting from this subjective value a similar computation made at the beginning of the period" (Edwards and Bell, 1973, 24–5). In contrast to historic cost asset valuation, this method assumes that the value is totally based on expectations about future income. Future cash flows are the basis of the measurement of both asset value and income. Income and value based on this method will be dependent on managers' subjective expectations, and consequently it generates subjective information output. As future receipts must be predicted and cannot, by definition, be measured, one is, under the Fisher tradition, measuring the manager's expectations or feelings

about the future. In the Fisher tradition we do not have knowledge, only beliefs.

The market value alternative provides another basis for determining income and value. Only in perfect markets will this approach represent an application of the economic value approach. The current market values are based on (usually imperfect) market judgements. Market values are social facts. However, the collective intentionality and perception that determine the value may be unstable, and thus the value may change. Recently we have seen in stock markets the institutionalized power of the instability of the market price. Market prices also include non-realized income, which may or may not be translated into cash benefits for the holder. The accounting stewardship model of measuring income and value is trapped between the conflicting intentions of producing objective accounting information founded on past transactions and being more economically credible by using subjective future-oriented information.

On this basis we may conclude that a historical financial statement establishes more knowledge than the Fisher and the market models. A historical financial statement is a statement reporting a true and fair view that sums up the financial situation of the company – i.e. it is a summation of all the phenomena or financial facts that *constitute* its financial situation. However, as some of the dimensions are future-oriented and some of the phenomena may be missing, it still has some proactive dimensions. The real truth depends not only upon trustworthy reporting practice but also upon trustworthy judgements of the business and institutional environments.

When we talk about plans and expectations regarding the future we have no objective facts about the future. However, integrated into the facts expressing the present situation is an array of possibilities, which may or may not become realized depending on a number of factors, one of which is what we decide to do. This is the proactive truth. This horizon of possibilities and impossibilities, of what we can do and what we cannot do, is always embedded in our factual information, no matter whether it is based on direct perception or indirect measurement. Since we cannot observe and measure all the relevant factors, our judgement as to the best line of action must, to some extent, be subjective. We try, however, in professional work to create epistemic platforms to assist our creation of plans for forecasting the future, as when we use our data to create plans and budgets, and when engineers and architects design their products.

The perception of reality introduced into this framework of pragmatic constructivism does not, as in realism, assume a mechanical world-view, in

which events can be predicted by a deterministic model. As we suggest that human values and logic are an important part of reality, and as values and real possibilities change over time and initiate developments, the implications are that some institutional phenomena change over time and that new ones arise. For planning and control we have institutional facts, and collective expectations as well as subjective expectations as to institutional phenomena. This distinction implies that some data used by accounting represent knowledge while some are only beliefs and assumptions. Necessarily, estimating future performance measurements will to some extent depend on the intuition of management, because not only past results but also the impact of future opportunities form part of the performance picture.

Proactive truth, real truth and learning

The subjectivity in our knowledge regarding values and intentions implies that any analysis of the situation and thus of the meaning of the data depends on the soundness of the operating logic, which is not limited to internal questions within a system but which concerns in particular the matching or coherency between several systems. The epistemological platform, in combination with a model reflecting the coherence of the performance system, is the proactive truth – i.e. the operating knowledge with which we conceive our situation in the world. It should never let us be totally blind as to what we can do. Our epistemological system can be used to describe the consequences of historical acts and decisions as well as estimating the consequences and feasibility of decisions and plans. Our facts are not blind to the future. A system connecting the elements of the dynamic life world with technological and economic perspectives is in place, and management can analyse the situation, make decisions and create plans according to financial and technological goals and operate successfully in the environment. The statements resulting from this system claim to be true, realizing the demands from phenomenology, coherence/logic and correspondence. We have analysed whether it fulfils the criteria of correspondence and the criteria of institutional, company and pragmatic coherence. The system set up in this way generates a *proactive truth*.

As measurements and analytical methods are always a reduction, we have to identify what we know in addition to measurements and our analytical language. At this level we are able to analyse the pragmatic coherence of our model at an integrated level. By making an overall evaluation of pragmatic

coherence we make judgements as to the extent to which the plan is operational and the financial evaluation sound. Does the business environment give reason to trust the valuation of our assets? What are the threats and risks to our business? Does the company have sufficient resources available, with future potential that can be managed in the company's sphere of interests? Can one, for each activity, develop inputs, processes and results in such a way that they are integrated and, together, can accomplish the overall goals of the company?

Our ability to judge pragmatic coherence, and thereby predict results, depends upon sufficient knowledge about both the means and the ends. Consequently, insufficient knowledge about means and ends reduces the possibilities for predicting and safeguarding results, and thereby also leads to higher uncertainty (Thompson, 1967). Measurements may not make sense if they are not interpreted in relation to feelings and more or less non-conscious knowledge about society. That involves another form of intuitive judgement. For example, an experienced and successful division manager questions his unit managers to ascertain *whether they believe in their plans or data and whether they lie awake at night because they are afraid of something*. In this way he judges whether the performance measurement model is valid with respect to being integrated into the company as a phenomenon – i.e. does it integrate facts, logic, value and language?

However, only the results of the activities show the degree to which the modelled company expresses the reality of the company. The *real truth* depends on the outcome of the accounting endeavours and may therefore differ from the proactive truth. Hereby the learning perspective becomes the objective of installing management accounting systems. The deviation is information about differences between the reality of the firm and the observed and controlled firm. Analysis of the control system should lead to improvements in the control system. The dynamics of the world will always tend to make proactive truth and real truth drift apart. The implemented learning process counters this problem.

Real truth is retrospective, and this is of no help in managing practice. It gives us a way to judge the level of truth of the proactive statements generated by the models, and the analysis as implemented through decisions and plans. Real truth is historical information only, however.

Depending on the quality of operationalization, one can observe the ability to create realizable plans and predict events with sufficient precision to enable the study of the degree to which proactive truth matches the real truth – i.e. the degree to which the outcome of activities matches the expected or planned

results. The observation and analysis of the difference or deviation between the expected and the realized is the vehicle of management control and accounting as a system of learning. Learning is the result of, and is based on, systematic analysis of the cause of the observed difference. The learning analysis looks for the reason for deviation in all the steps we have presented. The measurements may be misleading, the understanding of the institutional framework may be unsatisfactory, there may be hidden incoherencies, the abstractions may be too narrow or the phenomenological basis may be misleading, in that some new type of phenomenon may have occurred unnoticed. As a result of the analysis, the performance measurement system may be revised.

The learning process is necessary. This process differs, however, from Popper's (1965) idea of learning by creating bold hypotheses and trying to falsify them. Based on the proactive system, the company creates expectations and predictions (often these predictions are safe, but sometimes they are bold, as Popper recommends in his so-called "critical rationalism"). Here, however, we do not learn through attempts to falsify the predictions. The company learns from the marginal deviations. Falsification attempts would, generally, be disastrous. Learning must take place before falsification occurs. Learning in practice takes place in relation to any difference, any deviation. Attempts to learn by falsification work only when general laws are tested in experimental situations, in statistical studies or in analytical prototypes. When working with live cases in which the pragmatics of truth is essential, error elimination is attempted falsification. This implies, however, that survival depends on creating sufficient margin so that one can survive some degree of deviation; otherwise, no learning is possible, and if learning is impossible the firm's troubles will start to accumulate.

Conclusions and perspectives

Within the realm of accounting, realism is conventionally employed as a substantiation for the truthfulness of information produced. However, realism, when used in this way, is based on a misleading concept of reality. When you deal with human beings and companies you are dealing with intelligent systems, and not deterministic systems following an objective logic. Consequently, realism cannot solve the problems of information validity and truth but can only distort them. You have to observe the inner system and not just the outside system, as in the physical world, in order to understand and describe human reality.

We assume that social reality includes four integrated dimensions – facts, logic, human values and language – that are all important in constructing social reality (Nørreklit, Nørreklit and Israelsen, 2006). The application of this formulation of reality as a means of describing and explaining accounting practice ensures that legitimate claims of truthfulness can be made for accounting performance measurements, even though they are, inevitably, based heavily on professional judgement (as opposed to deterministic, real-world phenomena). In practice, these four dimensions can often provide a – more or less – integrated basis for explaining and justifying accounting behaviour; in current theorizing about the foundations of and in justifying accounting practice, however, they are missing. As accounting performance measurements are devised in practice, it is necessary to take all four of the dimensions into account in order to develop valid models. A model not integrating all the relevant dimensions cannot be expected to show valid results, and therefore the quality of truth may well be compromised by it.

This analysis also implies that we need an integrative theory of truth to underpin the truth of accounting performance measures. An integrative theory of truth draws on multiple aspects of truth: pragmatic, phenomenological, coherence/logic, as well as correspondence. Furthermore, it means that truth, for the accountant, can appropriately take the form of a learning process founded on the dynamic between proactive and real truth.

References

- Arbib, M.A., and Hesse, M. (1986). *The Construction of Reality*. Cambridge: Cambridge University Press.
- Arbnor, I., and Bjerke, B. (1997). *Methodology for Creating Business Knowledge*. London: Sage.
- Adorno, Th. W., Dahrendorf, R., and Pilot, H. (1991). *Der Positivismusstreit in der deutschen Soziologie*. Darmstadt: Luchterhand Literaturverlag.
- Argyris, C., and Kaplan, R. S. (1994). Implementing new knowledge: the case of activity-based costing. *Accounting Horizons*, 8(3), 83–105.
- Chambers, R. J. (1966). *Accounting Evaluation and Economic Behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
- de Haas, M., and Kleingeld, A. (1999). A multilevel design of performance measurement systems: enhancing strategic dialogue throughout the organization. *Management Accounting Research*, 10, 233–61.
- Descartes, R. (1988 [1628/9]). *Regulae ad directionem Ingenii*. In Cottingham, J., Stoothoff, R., and Murdoch, D. (eds.), *The Philosophical Writings of Descartes*. Cambridge: Cambridge University Press.

- Edelman, J. M. (1977). *Political Language: Words that Succeed and Politics that Fail*. New York: Academic Press.
- Edwards, E. O., and Bell, P. W. (1973). *The Theory and Measurement of Business Income*. Berkeley: University of California Press.
- Fisher, I. (1906). *The Nature of Capital and Income*. New York: Macmillan.
- (1930). *The Theory of Interest*. New York: Macmillan.
- Gadamer, H. G. (1960). *Warheit und Methode*. Tübingen: J. C. B. Mohr.
- Grady, (1991). Performance measurement: implementing strategy. *Management Accounting*, 72, 49–53.
- Habermas, J. (1990). *Moral Consciousness and Communicative Action*. Cambridge: Polity Press.
- Heidegger, M. (1927). *Sein und Zeit*. Halle: Max Niemeyer.
- Husserl, E. (1913). *Ideen zu einer reinen Phänomenologie und Phänomenologischen Philosophie*. The Hague: M. Nijhoff.
- Hicks, J. R. (1946). *Value and Capital*. London: Oxford University Press.
- Kant, I. (1999 [1787]). *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- Kaplan, R. S., and Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Leach, R. G., and Stamp, E. (1981). *British Accounting Standards: The First Ten Years*. Cambridge: Woodhead Faulkner.
- MacNeal, K. (1970 [1939]). *Truth in Accounting*. Lawrence, KS: Scholars Book Company.
- Mautz, R. A., and Sharaf, H. A. (1961). *The Philosophy of Auditing*. Sarasota, FL: American Accounting Association.
- May, G. O. (1972). *Financial Accounting: A Distillation of Experience*. Houston: Scholars Book Company.
- Meyer, J., and Rowan, B. (1977). Institutionalized organizations: formal structure of myth and ceremony. *American Journal of Sociology*, 83(2), 340–63.
- Nørreklit, H. (2000). The balance on the balanced scorecard: a critical analysis of some of its assumptions. *Management Accounting Research*, 11(1), 65–88.
- Nørreklit, H., and Schoenfeld, H. M. (2000). Controlling multinational companies: an attempt to analyze some unresolved issues. *International Journal of Accounting*, 35(3), 415–30.
- Nørreklit, L., Nørreklit, H., and Israelsen, P. (2006). Validity of management control topoi: towards a constructivist pragmatism. *Management Accounting Research*, 17(1), 42–71.
- Nørreklit, L., and Schoenfeld, H. M. (1996). *Resources of the Firm, Creating, Controlling and Accounting*. Copenhagen: DJØF Publishing.
- Paton, W. A. (1962). *Accounting Theory*. Houston: Scholars Book Company.
- Popper, K. R. (1965). *Conjectures and Refutations*. New York: Harper and Row.
- Porter, T. M. (1995). *Trust in Numbers*. Princeton, NJ: Princeton University Press.
- Searle, J. (1995). *The Construction of Social Reality*. Harmondsworth: Penguin.
- Solomons, D. (1991a). Accounting and social change: a neutralist view. *Accounting, Organizations and Society*, 16(3), 287–95.
- Stamp, E. (1985). The politics of professional accounting research: some personal reflections. *Accounting, Organizations and Society*, 10(1), 111–23.

- Stamp, E., and Marley, C. (1970). *Accounting Principles and the City Code: The Case for Reform*. London: Butterworths.
- Sterling, R.R. (1970). *Towards a Science of Accounting*. Lawrence, KS: Scholars Book Company.
- (1979). *Theory of the Measurement of Enterprise Income*. Lawrence: University Press of Kansas.
- Strauss, A., and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.
- Thomas, A.L. (1969). *The Allocation Problem in Financial Accounting Theory*, Studies in Accounting Research no. 3. Evanston, IL: American Accounting Association.
- (1974). *The Allocation Problem: Part Two*, Studies in Accounting Research no. 9. Sarasota, FL: American Accounting Association.
- Thompson, J.D. (1967). *Organizations in Action*. New York: McGraw-Hill.
- Tinker, A.M., Merino, B.D., and Neimark, M.D. (1982). The normative origins of positive theories: ideology and accounting thought. *Accounting, Organizations and Society*, 7(2), 167–200.
- Tinker, T. (1991). The accountant as partisan. *Accounting, Organizations and Society*, 16(3), 297–310.
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

The validity of measurement frameworks: measurement theory

Stephen Pike and Goran Roos

Introduction

When considering measurement, the first practical questions faced by managers should be: what is it that requires measurement, and what is measurement? Answering these two questions at something more than a superficial level can take managers a long way towards the design and development of a measurements system suited to their needs.

One surprising issue often overlooked when measuring systems are designed concerns the nature and purpose of the measurement. Quite often this is brutally exposed by the chief executive if, when presented with new results for the first time, he or she says “So what?”. A considerable benchmarking industry has grown up around the subject of performance measurement. Performance measurement assumes that the results it provides and the benchmarking activities that accompany it are useful in themselves. The motivation of the chief executive in the statement above is to question value – that is, how valuable this performance is to the company. Would it make a difference to any of his/her stakeholders or shareholders if the performance of certain factors were to increase by 10 per cent or if they were to decline by 10 per cent? The distinction between value and performance is critical, and, in general, performance measurement should be seen only as the first step on the way to providing useful value-based output. Measuring value adds some particular difficulties, especially with the definition of value, and involves the axiology of the individual.

Measurement is the process of assigning numbers to things in such a way that the relationships of the numbers reflect the relationships of the attributes of the things being measured. This definition of measurement applies not just to the simple and familiar measurements, such as distance and mass, but also to the measurement of complex things, such as the value of businesses that need diverse sets of attributes to describe them.

The relevant theory in this regard is measurement theory, a branch of applied mathematics. The fundamental idea of measurement theory is that a measurement is not the same as the object being measured but is, instead, a representation of it. This means that, if conclusions are to be drawn about the object, it is necessary to take into account the nature of the correspondence between the attribute and the measurement. Proper use of measurement theory greatly reduces the possibility of making meaningless statements; for example, although thirty is twice fifteen, a temperature of 30°C is not twice as hot as 15°C, since there is no simple correspondence between the numerical measure and the object.

Measurement, especially value measurement, is a surprisingly commonplace activity. Measurement is typically based either on proper measurement or, more usually, on subjective judgement. When the finance director is calculating the goodness of a proposal to invest in a new venture, or some new plant or equipment, he is using rigorous measurement, while a quick choice between two alternative products in a shop requires much less rigour, but the processes are surprisingly similar. Data, which may be hard or subjective, are collected and used with a measurement model, which may be formal and governed by hard rules or may be ad hoc. In both cases the measurement model is used to develop an answer, which then becomes part of the management decision or the choice between products. However, there are important distinctions between proper measurement systems and less rigorous approaches, which are termed “indicators”. The decision whether to embark on building a measurement system or to use a set of indicators depends on the situation faced now or likely to be faced in the future. Both measurements and indicators have their advantages and disadvantages, and table 10.1 sets them out.

The choice of route to follow depends on whether an accurate measurement is required upon which critical decisions may have to be taken, or whether a group of indicators that monitor changes as they occur, but cannot be used for decision making with safety, will suffice. If reliable information to make decisions about the future is required, then a proper measurement system must be used. If the monitoring of progress towards targets with moderate trust in the results is the requirement, then indicators may be used. In this chapter, it is assumed that the reader needs to measure accurately.

The principles used in the measurement of the financial resources of a company, or part of a company or meta-activity across a company are complete, sophisticated and well documented. They are not unique, as different financial jurisdictions have subtly different rules for descriptions

Table 10.1: Comparison of proper measurement and indicators

	Measurement system	Indicators
Advantages	Accurate if built properly Produces a complete view of the object Data can be disclosed Results can be benchmarked Can be the basis of derived measures Can be used with other business models Transparent and auditable Takes multiple views of value into account	Quick to build Easy to operate
Disadvantages	Takes care and time to set up Data requirement can be large Data quality requirements are stringent	Purpose-specific Cannot be benchmarked with safety Takes a single “average” view of value Cannot be built up to value complex objects Possibility of duplication

and combinations. A wide variety of schemes are available for recording intangible resources and have been reviewed by Pike and Roos (2004a), and there are a wide variety of justifications of such schemes as being right for one’s company needs. A few years ago the *Harvard Business Review* (Ittner and Larcker, 2003) published an excellent set of guides for the development of useful measurement systems for non-financial resource measurement. They list four mistakes common in business measurement systems.

- (1) Not connecting the measurements to strategy (or what really needs to be measured).
- (2) Not ensuring that there are causal links between the measure and the phenomena to be measured.
- (3) Not setting the right performance metrics and targets.
- (4) Measuring incorrectly.

To these a fifth might be added, which is that the measurement system needs to be as compact as possible. Many companies have more performance measurements than they use and even know of. Many measures are developed independently to serve specific purposes and many become

obsolete rapidly. Few are turned off. It is likely that most companies have considerable redundancy amongst their measurements and measurement schemes, and thus sponsor costly activities the uselessness of which is often very apparent to those close to the coalface.

(5) Not dealing with redundancy or unwieldy measurement systems.

When the purpose of measurement is to support the external publication of company performance, then there have to be some additional requirements on the measurement system. The further mistakes that are made by measurement systems are as follows.

(6) Not auditable (by an independent third party) and hence unreliable.

(7) Unable to generate the information needed by shareholders, investors or other relevant stakeholders.

A final impediment to good measurement concerns the size of measurement systems. If management is over-dependent on measurement, justifiable accusations of micro-management are levelled and unwanted behaviours tend to result. This latter point arises because people tend to want to improve performance, and tend to focus on many trivial elements in an over-elaborate measurement system. In doing this they lose sight of the bigger and more important picture.

In addition to Garvin's statement that "if you can't measure it, you can't manage it" (Garvin, 1993), there are a great many reasons why business or parts of businesses should be measured beyond what is required by existing statutes. In recent years the focus has been on the measurement of intellectual capital. The reasons for this may be considered as broadly falling into one of two categories. The first is that the valuation or evaluation will be useful in monitoring events, and especially non-financial events, inside the company. The second is that there may be reasons to disclose the non-financial results of an evaluation externally (Pike, Rylander and Roos, 2002). Indeed, supranational bodies such as the Organisation for Economic Co-operation and Development (OECD, 1999) and the European Union (Eustace, 2000), and regulatory and representative bodies such as the US Securities and Exchange Commission (SEC) (Turner, 2001), Federal Accounting Standards Board (FASB, 2001a, b) and the Institute of Chartered Accountants in England and Wales,¹ are actively sponsoring work concerning the disclosure of elements of organizational operations that are non-financial and fall into the field of intellectual capital measurement.

¹ See the ICAEW library reading list on intangibles, available at www.icaew.co.uk.

The two reasons for measurement, internal and external, are quite different in purpose and can lead to quite different forms of measurement (for further discussion of measurement in different situations, see M'Pherson and Pike, 2001, Fletcher *et al.*, 2003, Burgman and Roos, 2004, Pike and Roos, 2004b, Roos, Pike and Fernström, 2004, Gray, Roos and Rastas, 2004, and Roos and Jatur, 2005).

What should be measured?

When applied to businesses, this is a complex question, in two parts. In the first place, it requires boundaries to be placed around the “object” to be measured so that it is absolutely certain what it is that is to be measured, and, in the second place, it requires selection of the necessary aspects of the object to be measured.

The question about boundaries is a fundamental one and addresses issues such as whether it is necessary to measure the performance of a meta-activity – e.g. whether to view a knowledge management investment as an isolated entity or to measure the impact it has on the company. The measures required for the second case are, clearly, more extensive than the first and require a measurement of the whole company. If a company is seen as a system of interacting resources and processes, then modifications to one parameter have the potential to affect all the others. When considering major meta-activities, such as knowledge management or strategic human resource management (SHRM), then it should be obvious that the only way to assess the meta-activity is to consider its effects on the whole system. It is commonplace to consider the meta-activity in isolation, but to do so can lead only to indicators² concerning the intrinsic nature of the meta-activity. For example, a knowledge management system may be measured and found to have a certain number of documents within it, a certain number of discussion fora, and an information retrieval time of so many mega-bytes per second and so on. These may be benchmarked, but they retain an intrinsic quality; even with approaches, such as the balanced scorecard, that have an outcome

² An indicator is a roughly estimated representation of an object that may suffice for local needs but that is prone to errors, omissions and duplication. A measure is a numerical representation of an object in which all the attributes of the object are included in the representation in a manner compliant with measurement theory and all the measures and manipulations are also compliant with measurement theory.

element, the outcomes are simplistic and cannot measure the effect the meta-activity really has on the company as a whole. They do not measure the instrumental value of the meta-activity or its extrinsic value. The former requires a whole-company perspective, and the latter a perspective that reaches beyond the company to stakeholders such as analysts, customers and suppliers.

The question about what aspects of the entity to be measured should be included is an extension of the above arguments about boundaries. Just as the measurement of a meta-activity has to include the system(s) of which it is a part, the attributes to be measured must include all the attributes that any legitimate observer or stakeholder believes to constitute the entity to be measured. Partial measures carry obvious dangers, in that important contributory features could be omitted and no estimate can be given as to the seriousness of the omission. In extreme cases it can lead to the belief that the object being measured is performing well and delivering value when it is not – and is plainly not. To avoid this situation, it is usual to consult widely through the company and beyond it to canvass opinion of what the meta-activity is and does. In other words, all stakeholders have to be consulted to ensure that the boundaries of the meta-activity are set correctly and that all elements of it and what it does are included for measurement.

The need to select stakeholders with authoritative views about the company, meta-activity or object to be measured is very important. In order to measure a company, meta-activity or object, not only must the object be completely defined, so must the context of the measurement. A valuation of a company in the context of its performance in the market is quite a different matter from a valuation in the context of internal efficiency. In all measurements there is an implied comparison, or at least the ability to carry out a comparison. In the former case the implied comparison is with competing companies, and in the latter it is amongst management areas.

Stakeholders have to be chosen with care so that their views are authoritative and relevant to the context of the problem. A stakeholder is defined as anyone who can affect an outcome that is valued by the organization. The framework developed by Agle, Mitchell and Sonnenfield (1999) to nominate stakeholders is a useful starting point, and is shown in figure 10.1. The framework uses the criteria of power, legitimacy and urgency to categorize stakeholders, and goes on to categorize by importance, with definitive stakeholders in the centre as most important. Stakeholders on the periphery have fewer “rights of claim”.

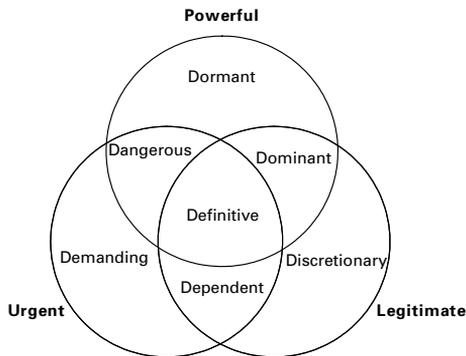


Figure 10.1: Stakeholder “claim” categorization

Previously, Mason and Mitroff (1981) had applied the concept of the stakeholder in strategic assumption surfacing and testing (SAST), a planning process for business strategic and policy planning. They stress the importance of identifying as many of the stakeholders as possible. Mitroff and Linstone (1993) use SAST in a case study of a drug company with ten groups of stakeholders involved in the company’s business. Mason and Mitroff (1981) also propose a range of stakeholder generation methods that help policy analysts identify stakeholders in various dimensions. They give seven categories of stakeholder, which may be contrasted with the seven in figure 10.1. Selecting the stakeholder identification method is a function of the nature of the stakeholder group’s likely constituents and the nature of the problem that is being addressed. The seven categories of Mason and Mitroff are as follows:

- (1) imperative;
- (2) positional;
- (3) reputational;
- (4) social participation;
- (5) opinion-leadership;
- (6) demographic; and
- (7) organizational.

Measurement

The formalization of measurement theory is a surprisingly recent development. This may seem surprising, since measurement has been taking place ever since herders needed to count livestock millennia ago. Measurement

theory can be traced back to Eudoxus of Cnidus.³ His work on the theory of proportions addressed important issues with the practical use of simple measurement; for example, it was impossible to compare numbers when they were irrational. The ideas of the modern theory of measurement date only from the nineteenth-century work of Helmholtz, and others (Helmholtz, 1887). Although most widely known for his work on electromagnetism, he began his academic career with physiology, optics and philosophy. In the latter part of his career he published works on counting and measurement, which laid the basis for one-dimensional extensive relational measurement theory.

The formalization of measurement theory belongs to the twentieth century. The primary motivation for this was the need to understand what it means to measure things in the social sciences. Many of the entities to be measured were not simple physical objects and were therefore hard to measure, much as intangible resources would appear to be today. The catalyst for the formalization of measurement theory is generally accepted to be the psychologist S. S. Stevens, with later interest from scientists in the field of quantum physics, although it was not until the 1970s that measurement was fully axiomatized (Stevens, 1946). This was accomplished with the publications of Suppes and others at Stanford University (Scott and Suppes, 1958; Suppes and Zines, 1963; Krantz *et al.*, 1971, 1989, 1990; Narens, 1985). They showed that numerical representations of values and laws are only numerical codes of algebraic structures representing the real properties of these values and laws. Thus, hierarchical structures are primary representations of values and laws.

According to Krantz *et al.*, the main propositions of measurement theory are as follows.

- (1) Numerical representations of quantities and laws of nature are determined by the set of axioms for corresponding empirical systems – algebraic systems with some sets of relations and operations.
- (2) These numerical representations are unique up to some sets of allowable transformations (such as a change of measurement units).
- (3) All physical attributes may be embedded into the structure of physical quantities.
- (4) Physical laws are simple, because of the procedure of simultaneous scaling of all attributes involved in the law (there is no machine learning method to perform such discovering of laws).

³ For a biography of Eudoxus, see Encyclopaedia Britannica. A useful review can be found at www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Eudoxus.html.

(5) The same axiomatic approach is also applicable not just for physical attributes and laws but for many other attributes from other domains (such as psychology), using polynomial and other representations.

Pfanzagl (1971) suggests that other, non-representational measurement approaches are oxymorons, and that measurement using the Likert scale is “measurement by fiat”. Such measurement theorists believe that representational or axiomatic measurement would allow psychology to replace measurement by fiat with more defensible measurement procedures (Krantz *et al.*, 1971, 1989, 1990).

While representational measurement theory is used by the scientific world, some, such as Dawes (1994) and Smith (Dawes and Smith, 1985), believe that axiomatic measurement theory has never really been accepted outside the world of the natural sciences. Dawes believes that the reason for this may lie in the difficulty of the underlying abstract mathematics, the lack of demonstrated experimental use and the problem of dealing with errors. Others, such as Marley (1992), believe that it is premature to pronounce representational measurement a failure in psychometrics, as axiomatic measurement has provided a well-founded framework to assist with the development of theories.

It is a debatable point, but it would not be unreasonable to see business measurement placed between the natural sciences and psychometrics, perhaps within social sciences in general. Certainly, it is a field that has generally been dominated by the measurement of financial and physical assets using procedures and the most basic of axiologies.

Before proceeding further, it is important to remember that performance measurement is an input and not the goal, as it says nothing about whether the result is important or not. This is undoubtedly a contributing factor in companies in which there has been an unrestrained growth in measurement; measurements are taken because they can be rather than because they are needed in a value measurement system. When measuring companies, value measurement should be undertaken.

Value measurement brings instant consequences, stemming from the fact that the only safe approach to axiology is to assume that different people have different views of value since value is dependent on personal values. It is often summarized with the statement “Beauty is in the eye of the beholder”. For example, employees, managers, analysts and regulators will have differing views of what is valuable in a company, but all the views must be taken into account, as failure to do so could mean the exclusion of some aspect that is of value or could lead to poor decisions based on an

incomplete picture. Four important outcomes from the independence of value definition must be borne in mind.

- (1) The object to be measured or valued and the context in which the object subsists must be precisely defined.
- (2) The definition is inclusive in its detail of all opinions and requirements from all stakeholders.
- (3) All participants (stakeholders) have equal dignity or importance, at least to begin with.
- (4) Every participant is accountable for the veracity of his/her position.

In classical texts, axiology is introduced, explained and exemplified in terms of simple objects, but a business, part of a business or meta-activity across a business is not a simple object. To deal with complex objects requires us to extend the principles of axiology, and the method easiest to operationalize invokes multi-attribute value theory (MAVT). Note that multi-attribute value theory is often considered to be similar to multi-attribute utility theory (MAUT) but with no uncertainties about the consequences of the alternatives, whereas MAUT explicitly recognizes uncertainties (Keeney and Raiffa, 1976).

MAVT allows the representation of complex entities using a hierarchical structure in which the elements of value are contained in a complete set of mutually independent attributes. Such value measurement structures can be made operational in conjoint structures by the incorporation of algorithms to represent the subjective judgements made by stakeholders. For reliable use, it requires the algorithms to be compliant with measurement theory in all places. The basic idea in conjoint measurement is to measure one attribute against another. Clearly, this must involve common scales, and one in which the scale ends have a defined meaning. In practice, making scales commensurable requires normalization onto a 0 to 1 scale and that input data are expressed on a ratio scale. When other scaling systems are used, such as interval and ordinal, Tversky, Slovic and Sattath (1988) have shown that they are incompatible with the axioms of measurement theory and commensurable scales cannot be constructed with them.

Constructing a practical measurement system for a complex object such as a business is a two-step process. In the first step, the object in its context is defined by the stakeholders, taking into strict account the legitimacy of the stakeholder and the implied comparison that the context will provide. A hierarchical measurements system is an ordered triple, a non-empty set A , containing all the attributes (a_i) of the entity, the relationships (r_i) between them and the operations (o_i) upon them. These are usually expressed as

$$A = \{a_i, r_i, o_i\} \quad \text{for } i = 1, n \text{ attributes}$$

For the set $A = \{a_i, r_i, o_i\}$ for $i = 1, n$, this means that for a set containing n elements

$$\sum a_i = A \quad \text{for } i = 1, n \text{ and}$$

$$a_i \cap a_j = 0 \quad \text{for } i, j = 1, n \text{ and } i \neq j$$

Regrettably, the only test that can be applied to demonstrate compliance with these conditions to prove this is proof by exhaustion.

The hierarchical structure constructed along these lines simply describes the object to be measured but nothing more. It shows what “ought” to be measured. The second step is to build an operational isomorph. The practical problem that is almost always encountered is that some of the attributes that “ought” to be measured cannot be easily measured in practice, and thus the stronger condition of homeomorphy cannot be invoked.

Thus, our description of the entity to be measured ($A = \{a_i, r_i, o_i\}$ for $i = 1, n$) appears as an isomorphic measurement structure in which the attributes a_i are replaced in part or in whole by proxies b_i . As the measurement structure is an isomorph, r_i and o_i are preserved and the measurement structure is represented by another ordered triple, B , where

$$B = \{b_i, r_i, o_i\} \quad \text{for } i = 1, n$$

Clearly, it is necessary to ensure that the proxies (which can be measured reliably and reproducibly) are acceptable. Since these are not exactly the same as the defined attributes that “ought” to be measured, it is necessary to test them to ensure that the conditions of completeness and distinctness have not been violated, but most of all that the aggregated meaning of B approximates to A in a way that is acceptable to the stakeholders. The method to test this is, again, proof by exhaustion, and the tests are as follows.

Assuming true isomorphism and $a_i \cong b_i$, i.e. $b \neq \{bb_k\}$, then

$$b_i \approx a_i \quad \text{and}$$

$$b_i \cap b_j = 0 \quad \text{for } i, j = 1, n \text{ and } i \neq j$$

It is easy to show and quantify the difference that sloppy mathematics makes to the results generated by a measurement system, but it is much harder to quantify the difference made by ill-chosen proxy measures.

It need hardly be said that ad hoc measurement systems, which often abound in companies and which are often ill-formed, are all dangerous.

The same is true for “indicators” – that is, less rigorous measurements the characteristics of which were given in table 10.1 – but, in their case, the approximations and inaccuracies are understood and accepted in advance.

Given an acceptable isomorphic measurement system \mathbf{B} , its measurable attributes are b_i with relationships r_i . All that is now needed is to consider the nature of the binary operations o_i . The simplest of aggregation algorithms is weighted addition, with aggregated value V of n attributes defined as

$$V = \sum_{i=1}^n w_i v_i \quad \text{and} \quad \text{where} \quad \sum_{i=1}^n w_i = 1$$

The simplicity of the weighted addition algorithm is often problematic, as it does not have the ability to show complex combination behaviours. This is especially important when the loss of performance of one combining attribute should lead to a complete loss of value in the combined higher-level attribute and this cannot be compensated for by a contribution from the other combining attribute. Marichal gives an excellent account of aggregation functions and their properties (Marichal, 1998).

When a measurement system is used, the results are dependent to a large degree on the nature of the combination algorithm. Clearly, it must be selected with care and conform to certain conditions of propositional logic. Failure to do this often introduces catastrophically large and variable errors into calculations. The key propositions that an algorithm must satisfy are those of commutativity and associativity, and they can be proved by algebraic means:

$$\begin{aligned} (f \circ g) \circ h &= f \circ (g \circ h) = f \circ g \circ h && \text{(associativity)} \\ f \circ g &= g \circ f && \text{(commutativity)} \end{aligned}$$

where \circ is the generalized binary operation of the aggregation function.

The penultimate step in constructing a measurement system is to customize copies of it so that it represents the behaviours of the individual stakeholders. This step ensures compliance with axiological requirements, in that the individual’s views are maintained without interference or the averaging of those results from consensus processes. In practice, this means asking for opinions on the relative importances of the attributes, the natures of the attribute combinations and the limits of performance. The last point is an interesting one: measurements consist of a number representing an amount and something that describes the scale – that is, the units.

In complex value measurement systems native performance scales are collapsed onto a non-dimensional value scale, which is normalized, between 0 and 1. The task that faces us is to define what 0 and 1 mean. It is usual to set 0 as that performance level that just becomes useful – in other words, “the threshold of uselessness”. The meaning of 1 has two common alternatives: either that it means the “best in class” or that it is some internally set strategic target. The choice between them is a matter for the company, but it is important that the basis be known.

The final step in producing a measurement system concerns the performance data to be used to operate the measurement system. All performance measures have two parts, the amount and the scale, but it is important to realize that there are many types of scale. If reliable results are to be obtained then it is important that data are collected on an appropriate scale. There are five scales, and they are shown in table 10.2.

For the purposes of proper measurement, only ratio or absolute scales are acceptable. To illustrate this, consider temperature. The Fahrenheit and Celsius scales are both interval scales, in that they do not have a meaningful zero and a transformation of the type $t(m) = c \cdot m + d$ is required to translate between them. The meaningful zero is the crucial point, and can lead to the absurd proposition that 30°C is twice as hot as 15°C . Where temperature is concerned, the only scale that is acceptable is the one used in the world of science: the Kelvin scale, with its zero being absolute zero.

Bearing in mind the multiplicity of units that may arise and the fact that a measurement system requires a meaningful zero to one scale, all data inputs must be commensurable, which, in practice, means normalizable. Some other conditions are also required of performance data to ensure that the measurement system functions properly. These are that

$$X \geq Y \quad \text{if and only if} \quad X + Z \geq Y + Z \quad (\text{monotonicity})$$

$$X + Y > X \quad (\text{positivity})$$

There exists a natural number n such that $nX \geq Y$

$$(\text{where } 1X = X \quad \text{and} \quad (n + 1)X = nX + X) \quad (\text{Archimedean condition})$$

In general, data can come from three sources. The first is from direct observation, the second is from a reliable simulation or business model and the third is by expert opinion. In all cases, the requirement for ratio or absolute scales applies. Of the three, expert opinion would seem to be the weakest, but it is quite valid if the expert is representing him/herself alone, such as if the expert were a customer in a customer-satisfaction-like survey.

Table 10.2: Description of scales

Name of scale	Typical description	Transformations	Allowed statistics
Nominal or categorical	A classification of the objects	Only those that preserve the fact that objects are different	Descriptive: frequencies, mode, information content; associative: chi-square
Ordinal	A ranking of the objects	Any monotonic increasing transformation, although a transformation that is not strictly increasing loses information	Descriptive: median, quantiles and quartiles; associative: Spearman's rank-order correlation coefficient, Kendall's tau, rho
Interval	Differences between values are meaningful, but not the values of the measure itself	Any affine transformation $t(m) = c^*m + d$, where c and d are constants; the origin and unit of measurement are arbitrary	As above, plus arithmetic mean, standard deviation
Ratio	There is a meaningful "zero" value and the ratios between values are meaningful	Any linear (similarity) transformation $t(m) = c^*m$, where c is a constant; the unit of measurement is arbitrary	As above, plus geometric mean
Absolute	All properties reflect the attribute	Only one-to-one transformations	All

If this is not the case then other measures have to be adopted to avoid the possibility that the expert may have a hidden agenda.

The use of measurement results

Inevitably, rigorous measurement can have the widest range of applications. Dependent on the original motivation for measurement, the results from well-executed value calculations can be used by themselves or in conjunction with other data (often financial data, but also other things), and can be used in simulating value over time with or without other data. Some basic application areas are shown below. The list is taken from work by Roos, Pike and Fernström (2006).

Value results alone

- (1) Value analysis of products, services or objects.
- (2) Value analysis of brands.
- (3) Value analysis of companies or parts of companies.
- (4) Value analysis of major functions in companies such as personnel and knowledge management.

Value results over time

- (5) Monitoring the progress of management initiatives.
- (6) Optimizing product or service replacement periods.

Value and financial results

- (7) Evaluating the value for money of candidate products or services.
- (8) Evaluating the value for money of candidate investments in infrastructure.
- (9) Evaluating the value for money of candidate investments in processes or services.
- (10) Benchmarking.
- (11) Disclosure to the market or to others.
- (12) Cost–benefit analyses.
- (13) Finding cost reduction options with greatest impact and least value penalty.

Value and financial results over time

- (14) Through life value-for-money assessment.
- (15) Evaluation of long-term strategic options.

Value and other results over time (example)

- (16) With risk to assess value at risk over time.

The sixteen points above are not exhaustive and should be seen as an illustrative list only. It should also be borne in mind that, in the preceding list, the word “value” had been taken to mean the single numerical outcome from the measurement hierarchy – in other words, the number at the top. There is, however, a wealth of other data available in the higher levels of a measurement hierarchy. It is common to find that about two levels of decomposition below the object/context lies a level wherein the functional tasks of managers appear. This correspondence disappears in the lower levels, since the hierarchy is constructed to rigorous requirements, whereas businesses are never organized with such rigour and separation of responsibility. Results can be taken from any level of aggregation in the hierarchy and used independently. If care is taken, they can also be used with one another.

References

- Agle, B., Mitchell, R., and Sonnenfeld, J. (1999). What matters to CEOs? An investigation into stakeholder attributes and salience, corporate performance and CEO values. *Academy of Management Journal*, 42(5), 507–25.
- Burgman, R., and Roos, G. (2004). Measuring, managing and delivering value performance in the public sector. *International Journal of Learning and Intellectual Capital*, 1(2), 132–49.
- Dawes, R. (1994). Psychological measurement. *Psychological Review*, 101, 278–81.
- Dawes, R., and Smith, T. (1985). Attitude and opinion measurement. In Lindzey, G., and Aronson, E. (eds.), *The Handbook of Social Psychology*, vol. I, 509–66. New York: Random House.
- Eustace, C. (2000). *The Intangible Economy: Impact and Policy Issues*, Report of the High-Level Expert Group on the Intangible Economy. Brussels: European Commission.
- FASB (2001a). *Business Combinations*, Statement no. 141. Washington, DC: Federal Accounting Standards Board.
- (2001b). *Goodwill and other Intangible Assets*, Statement no. 142. Washington, DC: Federal Accounting Standards Board.
- Fletcher, A., Guthrie, J., Steane, P., Roos, G., and Pike, S. (2003). Mapping stakeholder perceptions for a third sector organization. *Journal of Intellectual Capital*, 4(4), 505–27.
- Garvin, D. (1993). Building a learning organization. *Harvard Business Review*, 71(4), 78–91.
- Gray, D., Roos, G., and Rastas, T. (2004). What intangible resources do companies value, measure and report? A combination of UK and Finnish research. *International Journal of Learning and Intellectual Capital*, 1(3), 242–61.
- Helmholtz, H. (1887). Zahlen und Messen erkenntnistheoretisch betrachtet. In *Philosophische Aufsätze*, 17–52. Leipzig: Fues’s Verlag. (Trans. Bryan, C.L. (1930). *Counting and Measuring*. New York: Van Nostrand.)

- Ittner, C., and Larcker, D. (2003). Coming up short on nonfinancial performance measurement. *Harvard Business Review*, 81(11), 88–95.
- Keeney, R., and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley.
- Krantz, D., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement*, vol. I, *Additive and Polynomial Representations*. London: Academic Press.
- (1989). *Foundations of Measurement*, vol. II, *Geometrical, Threshold and Probabilistic Representations*. London: Academic Press.
- (1990). *Foundations of Measurement*, vol. III, *Representation, Axiomatization and Invariance*. London: Academic Press.
- M'Pherson, P., and Pike, S. (2001). Strategic analysis adapted for IC networks and Ebusiness networks. Paper presented at Twenty-First Annual International Conference of Strategic Management Society, San Francisco, 21–4 October.
- Marichal, J.-L. (1998). Aggregation operations for multicriteria decision aid. PhD thesis, University of Liège, Belgium.
- Marley, A. (1992). Measurement, models and autonomous agents. *Psychological Science*, 3, 93–6.
- Mason, R., and Mitroff, I. (1981). *Challenging Strategic Planning Assumptions: Theory, Cases and Techniques*. New York: John Wiley.
- Mitroff, I., and Linstone, H. (1993). *The Unbound Mind: Breaking the Chains of Traditional Business Thinking*. Oxford: Oxford University Press.
- Narens, L. (1985). *Abstract Measurement Theory*. Cambridge, MA: MIT Press.
- OECD (1999). *Measuring and Reporting Intellectual Capital: Experience, Issues and Prospects*, Proceedings of the International Symposium, Amsterdam, 9–11 June. Paris: Organisation for Economic Co-operation and Development.
- Pfanzagl, J. (1971). *Theory of Measurement* (in cooperation with Baumann, V., and Huber, H.), 2nd edn. Würzburg: Physica Verlag.
- Pike, S., and Roos, G. (2004a). Mathematics and modern business measurement. *Journal of Intellectual Capital*, 5(2), 243–56.
- (2004b). Measurement issues in intellectual capital: a review. Paper presented at International Intellectual Capital Conference, Taipei, 11–12 October.
- Pike, S., Rylander, A., and Roos, G. (2002). Intellectual capital management and disclosure. In Choo, C. W., and Bontis, N. (eds.), *The Strategic Management of Intellectual Capital and Organizational Knowledge*, 657–72. New York: Oxford University Press.
- Roos, G., and Jatur, S. (eds.) (2005). *Intellectual Capital, Management, Measurement and Disclosure*. Hyderabad: ICFAI University Press.
- Roos, G., Pike, S., and Fernström, L. (2004). Intellectual capital, management, measurement and disclosure. In Horváth, P., and Möller, K. (eds.), *Intangibles in der Unternehmenssteuerung: Strategien und Instrumente zur Wertsteigerung des immateriellen Kapitals*, 127–58. Munich: Verlag Frans Vahlen.
- (2006). *Managing Intellectual Capital in Practice*. New York: Butterworth Heinemann.
- Scott, D., and Suppes, P. (1958). Foundation aspects of theories of measurement. *Journal of Symbolic Logic*, 23, 113–28.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–80.

- Suppes, P., and Zines, J. (1963). Basic measurement theory. In Luce, R., Bush, R., and Galanter, E. (eds.), *Handbook of Mathematical Psychology*, vol. I, 1–76. New York: John Wiley.
- Turner, L. (2001). The times, they are a-changing. Speech given at Thirty-third Rocky Mountain Securities Conference, Colorado, 18 May. Available at www.sec.gov/news/speech/spch493.htm.
- Tversky, A., Slovic, P., and Sattath, S. (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371–84.

PART III

Performance measurement – practicalities and challenges

The first two parts of the book have concentrated largely on measurement theory, although many of the authors have drawn on their practical experiences as well as their research studies. The third part sees a change in tone, focusing primarily on measurement in practice, albeit measurement in *difficult* areas of practice.

The first contribution, from Tim Ambler and John Roberts, explores measurement from the customer perspective, introducing the concept of marketing dashboards and asking what they should contain. Ambler and Roberts explore the different measures that have been proposed over the years to track marketing performance, arguing that the search for a “silver metric” is somewhat futile. Instead, they call on marketers to adopt marketing dashboards that reflect the complexity of marketing accountability better.

The second contribution, from Andrew Likierman, explores risk and performance measurement. Likierman reviews the link between measurement and risk from three perspectives: in assessing organizational performance, in managing the organization and in reading the individual. He argues that performance measurement and risk are inextricably linked, on the grounds that discussing performance without reference to risk makes for meaningless comparisons.

The third contribution, from Rob Austin and Pat Larkey, moves firmly inside the organization, and asks how knowledge workers should be measured. This topic is particularly important in the developed economies. There, increasing numbers of firms are finding that they cannot compete on the basis of cost and therefore have to compete on the basis of value delivered. To create value, firms rely on their knowledge workers. Austin and Larkey explore the particular challenges associated with the measurement of

knowledge workers, arguing that they can be broadly categorized under the three headings of observability, motivation and TASK (talent, skill and knowledge) differentials. In reviewing the implications of these challenges, Austin and Larkey make an important distinction between measures for motivational purposes and measures for informational purposes, suggesting that the latter might be far more effective than the former for knowledge workers.

The fourth chapter relates to the themes raised by Austin and Larkey by exploring the question of how radical innovation might be measured. In her contribution, which is reproduced from the first edition of this book, Riitta Katila presents the results of an in-depth investigation into whether patent citation analysis can provide accurate insight into the impact of particular innovations. She finds that, far too often, the length of time over which patent citation studies are carried out is too short, and argues that the impact of many patents can only be seen some ten years after they were originally registered. As well as raising some important theoretical questions, Katila's study also raises some important practical issues. Her data show, for example, that collaboration can reduce the radicality of innovation outputs.

The final chapter in this part again explores measurement in difficult-to-measure contexts – this time from the perspective of project management. Graham Clark and Tony Powell present a novel measurement methodology – context-based measurement – that captures data on effort with associated contextual information. In their contribution, Clark and Powell illustrate not only the structure of context-based measurement but also its practical application, through a case study involving an engineering firm.

Choosing marketing dashboard metrics

Tim Ambler and John Roberts

Introduction

Marketing can be defined in many ways. Some see it as what the marketing department does, but different companies set different tasks for their specialist marketers and some have none at all. Others define marketing according to specific types of expenditure, such as advertising, promotion or market research, but this narrows “marketing” considerably, especially for firms that do not undertake those functions. Marketing has been defined for centuries – and well before marketing departments – as going to the marketplace with products for sale, selling them and returning with the resultant cash. Originally the term was used the other way about (taking cash and returning with products), but that is archaic, and this chapter defines “marketing” in this broad and traditional way – namely sourcing and harvesting cash flow as a result of pleasing customers and outdoing competitors (Ambler, 2003).

All organizations market their wares in this sense whether they recognize it or not. Not-for-profit organizations, such as charities, need inward cash flow just as much as traditional for-profit businesses. Marketing “performance” is, therefore, the extent to which the organization is succeeding with its short- and long-term inward cash goals compared with the resources being consumed in so doing. This may or may not have much to do with the marketing department, which, for example, may not be entrusted with pricing decisions or product development. This chapter therefore assumes that marketing is crucial and that its performance needs to be measured and monitored.

The productivity of marketing, in this broad view of inward cash compared with resources consumed, has become a hot topic (Bahadir and Tuli, 2002; Bruno, Parthasarathi and Singh, 2005; Debruyne and Hubbard, 2000; Morgan, Clark and Gooner, 2002; Rust, Lemon and Zeithaml, 2004), but protagonists seem polarized. One extreme argues that the ultimate purpose of marketing is to improve shareholder value and that marketing

performance should therefore be judged by a single, ultimate financial indicator. If shareholder value itself is not feasible, perhaps because it is confounded by too many other factors, then discounted cash flow (DCF) or return on marketing investment should be used. DCF is regularly reinvented and re-presented variously, as net present value, brand valuation (Perrier, 1997), customer lifetime value (Venkatesan and Kumar, 2004; Gupta and Lehmann, 2005) and customer equity (Rust, Lemon and Zeithaml, 2004). We therefore refer to all these as “DCF metrics”. Peppers and Rogers (2005) combine the ideas of DCF and ROI as “return on customer” (ROC).

The other extreme argues that any single bottom-line indicator (“silver metric”), attractive as such a simple solution may be, is impractical for a number of reasons. Performance is essentially multidimensional: superior performance against one objective cannot easily be traded off against lack of performance on another. Short- and long-term profits cannot be merged satisfactorily into a single number, because adequate short-term profit may be necessary for survival, irrespective of how attractive the prospect for long-term profit may appear. Customer, employee and societal imperatives cannot be discounted by larger and larger profits. Even for those focused only on net cash flow, the proxies for longer-term cash flow are non-financial in nature – e.g. customer satisfaction.

Firms have multiple goals and therefore need multiple performance measures. Moreover, even if a single number gave an adequate rating for the ultimate outcome, management needs to check the performance of each stage of the business model. Ideally, these multiple measures should be seen by senior management in a clear, integrated and concise package. Thus, management needs simultaneously to see a range of metrics on what is now known as the “dashboard”, a one-page or single-screen, easy-to-read summary of key marketing metrics (McGovern *et al.*, 2004; Reibstein *et al.*, 2005).

After some background comments concerning the gap between metrics theory and practice, we discuss the dashboard approach to relating and displaying multiple measures. Top managers use dashboards to help drive their businesses. They need to know performance to date, but also make choices for actions today to reach intended destinations. The dashboard of a modern car, for example, shows not only the amount of fuel remaining in the tank but also the miles per gallon since the car was last refuelled and the number of miles remaining available. While these two sets of measures are both related to performance, one past and one future, they fulfil different functions and may not always coincide.

The next section acknowledges DCF's suitability for deciding between marketing choices – i.e. planning – but questions the degree to which DCF metrics are appropriate for evaluating performance to date. We then discuss whether popular financial ratios, such as ROI and ROC, are useful metrics at all. The chapter then moves on to look at how financial and non-financial indicators should be chosen, and presents a model for assessing their relevance for the business. After considering the limitations of our approach, we outline where further research is needed and draw conclusions.

Metrics theory and practice

In line with the Marketing Science Institute, the top US industry–academic liaison body, raising metrics to its highest priority (e.g. Marketing Science Institute, 2000), recognition of the importance of, and the need for, accountability has grown sharply (e.g. Doyle, 2000; Moorman and Rust, 1999; Schultz, 2000; Shaw and Merrick, 2005; Srivastava, Shervani and Fahey, 1998). Marketers recognize the importance of performance measurement but progress has not been as rapid as some would have wished (such as the CMO Council, consisting primarily of digital technology companies, and the Marketing Leadership Council). One reason may be the indirect link between performance measurement and performance enhancement. Another may be that the lead times for changing marketing metrics systems exceed the expectancy for chief marketing officer job tenure, taken together with a lack of resources available for metrics, typically 1 per cent of the marketing budget (Glazier, Nelson and O'Sullivan, 2004).

A further explanation lies in the confusion between marketing (which the whole firm does) and the defined role of specialist marketers, which varies from company to company. For example, some have responsibility for pricing but others do not. Some CMOs have responsibility for sales and customer relationship management but others are responsible only for communications, promotions and market research. The metrics for a marketing dashboard are those for the whole company, whereas the metrics for assessing individual marketers may cover only those matters for which they are accountable.

Perhaps firms should transfer responsibility for gathering and reporting market data and performance to the financial function (Ambler, 2003). The typical response is that this may be logical but it would be too dangerous to give this licence to unsympathetic accountants. While it runs the dangers of

giving the measurement of non-financial performance to those trained only in counting in dollar metrics and of reducing the immersion of marketing managers in the market measurements that they are driving, it does have the dual advantages of perceived objectivity in gauging performance and freeing the marketing function to pursue rather than manipulate market outcomes.

Olsen (2005) may have supplied part of the answer, at least in respect of advertising. He researched evaluation with agencies and their clients and found, perhaps surprisingly, that clients rated ad evaluation as highly as creativity. In his survey, 82 per cent of advertisers (47 per cent highly, 35 per cent slightly) agreed that evaluation is as important as creative work. The explanation was that 96 per cent (83 per cent highly) believed it made budget justification easier in future. The CMO Council, however, rejects the notion that marketing accountability is motivated by a need to justify or increase budgets. No doubt some are and some are not. However, whatever the motivation, it does seem self-evident that access by CMOs to the resources necessary to maintain and grow marketing assets is facilitated by measurement of the effectiveness of how those funds will be applied, particularly if that measurement is perceived to be objective.

The American Marketing Association defines marketing accountability like this:

The responsibility for the systematic management of marketing resources and processes to achieve measurable gains in return on marketing investment and increased marketing efficiency, while maintaining quality and increasing the value of the corporation. (AMA, 2005, 1)

The significant part of that definition is the duality of short-term gains and enhancing the quality and value of the corporation – i.e. brand equity, which we define to be the intangible asset created by good marketing. Evaluating marketing performance requires both to be considered.

The dashboard approach: multiple measures

Firms have multiple goals and, since each goal needs at least one measure of progress towards it, the evaluation of marketing requires multiple performance metrics. Short-term profits and cash flow need to be adequate to ensure longer-term survival but, beyond that, firms have some flexibility in choosing whether to maximize short-term cash flow or invest some of it to improve future returns. Discounting cash flow is a valuable technique but in

reality the future, discounted by appropriate rates of interest, is not identical to the present. Cash in the bank is not the same as the proverbial cheque in the post.

The issue of whether a discounted future forecast can be equivalent to an actual present indicator is bigger than just whether future money can be equated to present money. Collins and Porras (1994) have been widely cited as showing that a focus on the bottom line is not just inadequate, it may be harmful. Employees are not motivated by working for money machines; they want to believe that their company is doing good. Johnson and Johnson is cited as such a company: yes, it makes money, but that is a by-product of meeting the needs of its products' end-users.

The firm's business model connects the actions the firm takes – e.g. the marketing mix – with the final outcomes – e.g. shareholder value. To drive the business, management needs to understand how the linkages work, or at least be alerted when something goes wrong. Thus, a dashboard may operate at two levels: the measures always in view, and the diagnostics that appear only when performance strays too far from preset controls.

Almost every company is seeking different goals in different ways. A company can succeed by being better at the same things or by being different. Neither is easy, and most firms try both. The same endeavours as other firms would justify the same, or similar, performance metrics. Differing endeavours each need their own metrics. Taken together, the firm needs enough measures to monitor its crucial marketing activities.

The question of how many measures are “enough” is similar to asking the length of the proverbial piece of string. At one extreme, senior managers, and especially those unsympathetic to marketing, whether narrowly or broadly – as here – defined, would like a single financial indicator, or, failing that, an index (e.g. Reichheld, 2003). The idea is to keep things simple by weighting all the detailed marketing metrics and combining them into a single figure used for top management control purposes. At the other extreme, control-oriented managers may like to have all the key numbers, and that, in a large and complex business, may be hundreds.

One flaw in the single index, or silver metric, concept can be illustrated by reference to our own health. If our blood pressure was very healthy but our cholesterol levels were high, a single index averaging the two would not be helpful. Brand health is like human health, in that what we are really looking for are the key indicators of *ill* health, which show where our attention is needed if we are to enjoy a long future, as well as opportunities to excel in new areas.

The flaw in the other extreme, a multiplicity of measures, is the limit to the number of different pieces of information we can consider simultaneously (Simon, 1955). Moreover, each metric multiplies because of the comparative figures needed to give it meaning. Performance can be compared to plans or to previous years, or to the market as a whole or to specific competitors. Looking forward, the likely performance and risks of alternative strategies can be compared with each other and with other ways the firm could spend the money.

A single metric in isolation is almost meaningless. We need to know how it has changed in the short and long terms, and how that compares with the planned outcome and competition. We may need to know the metric, and its comparatives, for the organization as a whole, and also the main brands or subunits. Indeed, it may not be sensible to combine metrics for dissimilar geographic regions or brands or units. Whether we are combining units or types of metrics, we should only aggregate like with like. Even if one started with a very few basic performance metrics, these comparisons and subunits could quickly multiply reporting to more pages, or screens, than the human brain can comprehend.

The dashboard concept starts from that point: never mind how many or how few metrics are theoretically justifiable, we should limit ourselves to making the best use of a single page or screen, which enables the viewer visually to link different pieces of information. The number of dials the modern passenger plane pilot has to monitor may actually have declined while the sophistication of the aircraft control systems has increased. Aircraft designers have recognized the need to keep pilots focused on the key instruments. When a problem arises, a warning sign lights up. Then further information can be sought, but such extended search will be highly focused.

Thus, the dashboard should include not only data to evaluate the progress and future direction of the firm but also the key indicators needed to hold the marketing function – when there is a separate marketing function – accountable on a regular basis. In addition to a “lighting up” facility for diagnostics when actual performance strays beyond preset controls, management needs to be able to drill down from these high-level metrics to the data used by junior marketing management. This can be called a clear “line of sight”. The cohesion of a firm is improved when all levels and functions are using common information embodied in a shared database. The interrelation between the board’s dashboard and other marketing information systems in the organization can be critical. Without it, the translation

between different languages leads inevitably to miscommunication, whether accidental or purposive.

Finally, the dashboard needs to reflect the dynamism of the firm within the turbulence of its market. If both the firm and its market are changing slowly, then the metrics will rarely need to be changed. If the firm or its market is changing rapidly and radically, however, the metrics may also need to change. This presents a problem: comparisons require continuity of metrics, and dynamism requires change. Firms can deal with that by double-running both old and new for the transitional period. Alternatively, core metrics can be monitored in every reporting time period, with others appearing only irregularly.

Should future DCF metrics be used for performance evaluation?

Diageo, the world's largest wine and spirit business, decided in 2004 that using metrics to measure past performance was not especially interesting, so it converted its dashboard into more intensive use of metrics as part of the forward planning and plan approval process.

This future orientation has merit: the past cannot be changed and is useful only to the extent that its review can improve future performance. For that reason, the view that "marketing is an investment and, unless you can measure its impact, your money is wasted" is wrong. By the time one can measure its impact, the success or waste of the investment has already taken place. In other words, subsequent measurement does not change history, although the way it is measured may change our perception of history – or, in this case, performance. Thus, measurement after the event does not change the event. We can learn from the investment and the results for the future but we cannot change what has already happened.

Furthermore, estimating the future profit from the investment does not alter the performance either – i.e. assure against waste. If it seems that the money would be wasted, the investment would presumably not take place. If it did take place and no profit ensued, measurement would not have prevented it.

The suitability of any tool, be it a spade or a market metric, depends on its intended usage. When we are looking forward to determine which marketing strategy will perform best, estimating the likely consequential cash flow, and risks, of each strategy is surely good practice. Discounting those cash flows back to NPV, whether in the guise of CLV, customer equity or brand

valuation, aids comparison. Moreover, the contextual variables, such as interest rates or economic growth, can be standardized across the alternative strategies to highlight the differences arising from the managerial variables. Better still, a range of scenarios can be applied consistently across all alternatives.

When we are evaluating performance to date, however, we have to use the real contextual variables, which are unlikely to match those in the plan. Variances can be analysed between those arising from the contextual changes and those from unexpected managerial variable performance, but such discussion quickly deteriorates once the accounts point out that management should have predicted the contextual variables better.

The problems with using future DCF, in whatever guise, for evaluating past performance to date go deeper than that. Let us explore two thought experiments. Suppose we have perfect vision of future cash flows and that the CMO has performed so badly to date that he is replaced at the beginning of the first future period by a brilliant performer. If the poor cash flow for the last period is compounded with the now excellent future performance, the overall DCF will exceed that of a merely average CMO. In other words, we have just persuaded ourselves that the bad performance we are evaluating is really rather good. Bring back the dreadful CMO. Discounted cash flows have meaning only within a very specific context of both actions within the firm and environmental factors affecting it.

One flaw in using future DCF for to-date evaluation is that we are taking credit for future performance not yet achieved. There are an infinite number of possible futures. Which one would you like to pick? For example, how can the future changes in the customer base be assessed? Which customers will be acquired and lost, with what impacts on cash flow? In businesses with high churn rates, such variation would be especially difficult to assess, not least because future churn depends on future marketing actions.

Assume that a characteristic of poor – or, at least, inexperienced – marketers is that they have inflated expectations of the results of their marketing. In this case, poor marketers will present higher DCFs than their more talented and experienced peers. This demonstrates at a bare minimum the need for some objective standards in judging the sanity of such estimates, given the moral hazard associated with their generation.

One reason why this future DCF topic arises for to-date evaluation is brand valuation or customer equity. In a single brand company these are the same thing: NPV or DCF of future net cash flows. The evaluators agree that short-term net cash flow, combined with the change in the marketing

asset, is the way to go, and so far so good. Then the marketing asset is quantified using the DCF technique, however, which takes credit now for future actions and cash flows. So we recognize that it is attractive to combine these two stages, because we agree that brand equity predicts future cash flows. Nonetheless it is not the only reason for future cash flows. Using future cash flows to determine performance to date is, therefore, flawed.

There is, however, an opposing school of thought that should be considered, and for which we have some sympathy. Whatever the theory, managers do, and probably should, try to combine past and future measures when making marketing decisions. The (relatively) known cash flow of the past (which cannot be influenced) differs from the more uncertain and only partially controllable cash flows of the future. Both are important and real. A typical problem is “Is past underperformance best addressed by sacking the manager, or should we retain the existing team to use the experience better?”. In looking back and forwards we need common metrics, or, at least, a transformation between the past and the future.

Another example of the value of future forecasts to assess performance to date is a yacht race. If two identical yachts leave the same port, headed for New York, we would accept that, if one yacht can reach New York in a day but the other will need three days, then the nearer yacht has performed better to date. We are using our expectation of future performance, namely getting to New York first, to assess performance to date. This example is illusory, however, because the future estimate is actually based on their present position: we are not estimating where the yachts are based on working back from the future but on knowing where they are now. We estimate how long each will take, from its current relative latitude and longitude. In other words, our assessment is based on present data.

The time factor becomes of major importance when we look at the differences between assessing performance to date and making better future choices (planning). The issue here is the relevance of DCF techniques that appear, as noted above, in many guises: brand valuation, CLV, customer equity and shareholder value where this is calculated on a DCF basis, as distinct from cash flow or profit to date plus change in share value to date.

Clearly, when we are looking forward, the financial outcomes can be compared year by year and discounted back to present value. Other things being equal, the strategy with the highest present value should be chosen, but that frequently ignores differing risks and the impact of future options. In

theory, DCF analysis can use different discount rates for different cash flows, but in practice it is hard. A strategy that delivers an apparently higher return but without flexibility for later adaptation to changing circumstances may not be as attractive as a strategy with a lower forecast return but flexibility. In theory, DCF can be made more sophisticated with the addition of decision trees and Bayesian statistical techniques. In other words, all possible decisions and outcomes can be identified and evaluated, but this rapidly becomes too complex for the comparison of immediate alternatives. In practice, the DCF calculations are not that subtle, and usually exclude risk analysis and possible future management decisions. They are normally predicated on one, very narrow, planning scenario. One can argue that it is simpler to use short-term asset outcomes – i.e. whether brand equity is strengthening or weakening – alongside cash flow.

In summary, DCF metrics may have validity for planning purposes, although the matter is not as clear-cut as it may seem. We cannot measure the future. For evaluating performance to date, however, DCF metrics are distinctly questionable in many of the applications to which they are put.

Return on investment

Return on investment was devised for comparing capital projects in which the investment is made once and the returns flow during the following years. ROI is the net return divided by the investment, or, more correctly, the incremental profit as a ratio of the incremental expenditure. Although, in theory, this can be calculated as a DCF stream of incremental cash flow versus the once-off marketing expenditure, in practice estimating the differential cash flows beyond the present year is too difficult, and ROI is calculated using the annual profit and expenditure either in whole or the incremental portions.

The first problem is whether marketing expenditure is an “investment” in the original sense. For capital projects, cash flows for some years and the once-off investment will typically end with some residual value. ROI is usually expected to be well above alternative potential investments, such as a bank deposit, but not to pay back in a year or two. Marketing expenditure, however, may influence later periods, but it is mostly directed to the current year. It has to be continued from year to year, maintaining the brand and the bottom line. While marketing expenditure may be partly investment for future years, most of it is current expense and all of it is

accounted for in the profit and loss account, not the balance sheet. While we must estimate the long-term effects of our marketing activity, otherwise we will underspend, it is important that the tail does not wag the dog in our accounting.

The second problem with ROI is that it requires the profit to be *divided* by expenditure, whereas all other bottom-line performance measures consider profit or cash flow after *deducting* expenditure. Division rather than subtraction distorts the metric. The profit, or economic value added or increase in shareholder value from marketing all require the costs to be deducted from sales revenue along with the other costs. Accounting and finance texts suggest that as long as capital is not constrained, residual income rather than ROI is a more appropriate measure (e.g. Peirson and Ramsay, 1998).

Box 11.1: Example of problems with ROI

The Brown Bread company is trading comfortably and is looking for modest growth. As the new finance director is asking the marketing department to maximize ROI, they produce the three plans in the table below. The net profit is given by the sales revenue after deducting the marketing spend and variable and fixed costs. So far as they can judge, and the numbers are not challenged by their colleagues, the first maximizes profit and cash flow and the second significantly enhances ROI. A third option maximizes ROI, in the sense that, with no marketing expenditure at all, ROI is infinite. Which should they choose? The last row treats plan C as the base case, since it has no marketing expenditure and looks at incremental costs and returns relative to that.

	This year	Plan A	Plan B	Plan C
Sales volume	100	110	93	80
Sales revenue	4,000	4,400	3,720	3,200
Marketing spend	800	950	700	0
Variable costs	1600	1760	1488	1280
Fixed costs	960	960	960	960
Net profit	640	730	572	960
Total ROI (%)	80	77	82	∞
Incremental ROI (%)		-24.2	-55.4	0

Plan B puts the sales into decline after some years of increase with the probability of a lower base in future years. Profit declines significantly but ROI increases by five percentage points. Greater cuts of marketing expenditure increase short-term profits considerably and also increase ROI. Let's assume that one of these scenarios has happened. On the incremental basis, using plan C as the baseline, plan A does at least perform less badly than plan B, but plan C still looks to be the best option.

The third problem with ROI is that pursuing it causes underperformance and suboptimal levels of activity. This arises from the law of diminishing returns. After the point of the profit response curve at which ROI is maximized, further sales will typically still make a profit, but at a diminishing rate until the curve crosses the X axis (the Y axis being profit and the X axis being marketing expenditure). There are exceptions but they are rare. For example, a seller of ice creams on a beach may find 30 per cent of those present to be on no-ice-cream diets and the remaining 70 per cent happy to buy one each but no more. If his marketing costs are low, then both profit and ROI are maximized at 70 per cent penetration, since a ceiling has been reached. In general, so long as the incremental profit exceeds the cost, total profit continues to increase but ROI progressively reduces. In technical terms, so long as the response function of sales to marketing expenses is continuous, maximizing ROI will not maximize profits.

The fourth problem with ROI is that the baseline – i.e. what would have happened without the expenditure? – is hard to determine, and mostly likely to be subjective judgement and able to be manipulated by the marketer.

The fifth problem with ROI is that it has become a fashionable term, the “new black”, for marketing productivity, and is used to describe any type of profit arising from marketing activities. As the director general of the (UK) Institute of Public Relations observed, “Ask 10 PRs to define ROI and you’ll get 10 different answers” (Farrington, 2004). The US American Marketing Association’s White Paper (AMA, 2005, 8) on marketing accountability identifies six “ROI measures currently used”:

- incremental sales revenue;
- ratio of cost to revenue;
- cost per sale generated;
- changes of financial value of sales generated;
- cost of new customer; and
- cost of old customer retention.

Not one of these six is actually ROI, although the second one is not far away, and thus the fifth problem is that marketers rarely mean “ROI” when they say “ROI”. Of course, this is a problem with usage rather than the ratio itself, but if usage is so confused that the metric has no consistent meaning then the metric, ROI in this case, should be abandoned.

The sixth, and possibly most serious, problem is that, in practice, ROI ignores the effect on the marketing asset (brand equity) and the longer term (which we take to be the same issue). In theory it does not have to, but estimating brand equity valuations into the future is beyond current skills,

although return on customer (see below) is a step in that direction. If marketing activities have generated £1 million in extra profit, after marketing costs of £0.5 million, ROI enthusiasts would applaud, especially if no other “investment would have paid back so handsomely”. If, however, the marketing activities had reduced the value of the marketing asset by £2 million, the story is reversed. This example underlines the importance, noted earlier, of using two types of measure to assess marketing performance: short-term profit or cash flow, and the change in brand equity. In practice, ROI assesses only one.

In summary, ROI metrics promote underperformance and short-termism and should not be used in marketing.

Return on customer

The quest for the silver metric to encapsulate marketing performance resembles the Dark Ages search for the unicorn. Sightings prove illusory. Peppers and Rogers (2005), for example, claim that maximizing ROC also maximizes both current-period and future profits. The concept has been launched with no little hype. Larry Kudlow, host of CNBC’s *Kudlow and Company*, has offered highest praise for return on customer:

Finally! A business metric that can drive better management and a higher stock price. I predict soon you’ll be hard pressed to find a company that isn’t tracking ROCsm.¹

The Peppers and Rogers definition is that “ROC equals a firm’s current-period cash flow from its customers plus any changes in the underlying customer equity, divided by the total customer equity at the beginning of the period” (Peppers and Rogers, 2005, 16). Adding together the change in short-term cash flow and the change in the marketing asset is valid and corresponds with our own view of performance measurement. The questions therefore become how the change in the marketing asset is measured and whether it is right to use a single (silver) measure of it rather than many. Customer equity is, in line with Rust, Lemon and Zeithaml (2004), taken to be the NPV of future cash flows, or DCF. Of course, one practical problem is to know who all the future customers will be, and then what cash flows will be contributed in response to the infinite permutation of marketing activities that the firm may undertake in future.

¹ See www.returnoncustomer.com.

In other words, the first problem with using this silver metric to evaluate performance is that it is most dependent on crystal-balling the immeasurable future. As Peppers and Rogers themselves concede in a slightly different context, “No one really knows what any company’s discounted cash flow is going to be in the future” (Peppers and Rogers, 2005, 19). They have failed to make the connection that DCF and customer equity are different labels for the same thing.

ROC is another matter. Unfortunately, teasing the meaning from the Peppers and Rogers formula requires the algebra that the editor has banished to the appendix. The next few paragraphs give a plain-language explanation as best we can, but please refer to the appendix for a more rigorous analysis.

If one replaces the current year’s customer equity (DCF) calculation by the equivalent calculation made a year earlier, the top line of the formula becomes the difference between the actual and forecast cash generated for the period under review, plus the difference between the two forecasts for the same outward years. Thus, if the forecasts are accurate, the return is zero. In other words, this ROC formula does not measure return on the value of the marketing assets so much as the *excess* of the return for the current period (compared to forecast) plus any *increase* in forecast, both taken as a ratio of customer equity.

What this means is that ROC is positive when we are doing better than was previously expected. This may be both sensible and practical. On the other hand, ROC just measures the *accuracy* of the forecast for the period just completed and the *consistency* of out years rather than marketing performance, which was the object of the exercise. Those who forecast low before and high now will appear better than those who did the reverse. In actual performance terms, either one could have been more successful than the other.

This relative performance aspect of ROC indicates that it will be particularly suspect to “gaming” – i.e. low budgeting and/or fattening short-term cash flow at the expense of the longer term whilst maintaining the high forecasts for the out years. This is a problem for all DCF techniques, but especially so for ROC. Competitor performance would be a useful yardstick, but that is not considered by ROC.

We also need to consider differences arising from year-to-year changes in the discount rate used. Increasing customer equity (DCF) may be due to real changes in future marketing prospects based on good work in the period, or improved prospects that have nothing to do with improved performance to date, or changes in discount rates or changes in environmental factors. Put another way, we are subject to all the problems of using

DCF as a performance indicator mentioned above. The technical interest rates and year shifts are not what we are examining. Worse still, the new forecast may have changed due to external factors, such as the political situation, which have nothing to do with the marketing performance being evaluated.

ROC brings DCF and ROI together but does not appear to provide a silver metric for assessing marketing performance. DCF can be a valid technique for comparing alternative future marketing activities, partly because they are simultaneous and therefore the external factors are the same for both. Time-shifting DCF calculations makes them less valuable for comparison because other factors are much less likely to be identical.

In summary, ROC measures the current discount rate used by the firm, which is hardly useful, and, beyond that, the *accuracy* of last year's forecast of cash flow in the period just ended, together with the *consistency* of the other two sets of forecasts across the two forecasting dates (last year and this year). In other words, the sceptic would see ROC as measuring forecasts rather than performance, but the supporter would see ROC as measuring performance relative to forecast.

Choosing financial and non-financial indicators

Perhaps the most surprising conclusion to be drawn from all this is the importance of understanding the difference between metrics for performance evaluation and planning. The “learning” stage – i.e. using performance evaluation to develop the business model – separates the two. Yes, performance should be evaluated by short-term cash flow (or profit) and the change in the brand equity, and, yes, brand equity can be financially valued using DCF, but combining these two stages – and ROC is one example – leads to serious problems. Brand equity is multidimensional and reducing it to a single measure loses diagnostic value. In the yacht example above, the remaining time is such an indicator, but to use it for management purposes the present location is crucial. We would not rely on the claimed distance to port any more than one should rely on a brand manager's claimed future cash flows.

Thus, metrics for evaluating performance to date and future alternative strategies may overlap. It is convenient they should for future evaluation, but the purposes, and thus the metrics, will differ. The three key stages are, therefore, as follows.

- (1) Using a range of financial and non-financial indicators for assessing the change in brand equity for performance evaluation – i.e. the short-term cash flow or profit performance to date plus indicators of significant changes to date in brand equity (evaluation).
- (2) Comparing what happened with what was expected to happen and adapting the firm's business model (learning).
- (3) Understanding alternative means of converting those indicators, together with new marketing resources, into future cash flow (planning).

A company should therefore combine the two sets of metrics (forward and back) on the same dashboard. There is a natural logical sequence to this. An examination of past performance should inform the consideration and evaluation of future plans. The choice of individual measures does not matter for our purposes here but would matter for those engaged in the practical business of supplying dashboards to C-suites (the executive level).

And there are more to dashboards than simply evaluating performance and planning. Other purposes include the following.

- (1) A better understanding of the firm's business model: what connects to what and with what results.
- (2) A better understanding of how its markets operate, and how it can better focus its resources to derive more performance from a given level of resource.
- (3) Through this better understanding of the external environment, the firm can pre-empt or react to changes in the customer, competitive or channel space.
- (4) Alignment within the firm through the communication of, and motivating commitment to, the key metrics (internal marketing).
- (5) Focus of the key leverage points that will determine whether the firm achieves its objectives or not.

The first can be illustrated by the way that promotions are sometimes monitored. Promotional "success" is sometimes measured by the height of the sales "spike" above the baseline. On the surface, the greater the sales on promotion, relative to sales without promotion, the more the promotion has achieved its objective. However, continual promotions educate consumers to wait for the next, and may well be harming brand equity (Mela, Gupta and Lehmann, 1997). Strong brands are indicated by not requiring (price) promotion. Moreover, promotions may well be damaging profits and shareholder value. Therefore, performance metrics need to align marketing activity with corporate goals; and herein lies the tension

between the complexity of a comprehensive analysis and the superficiality, if not misleading nature, of simplistic measures.

Other considerations when determining which metrics to use include the following.

- (1) **Diagnosticity:** a metric should indicate what action can be taken to improve performance.
- (2) **Redundancy:** metrics that move together can be reduced to whichever metric is the best, taking other considerations into account.
- (3) **Cost and availability:** knowing what parts of the business model one wants to monitor does not mean one can afford to do so, or even that the data are available at all. Internal data – e.g. sales – should be accessible but market data (customers and competitors, for instance) may not be. Systems put in place to collect them must have practical ways of being updated, otherwise marketing metrics full of hope and promise will be condemned to disuse and irrelevance.
- (4) **Reliability:** accountants sometimes dismiss non-financial indicators as flaky (Kallapur and Kwan, 2004). Increasing intention to purchase, as measured by market research, may be followed by declining sales (Morwitz, Steckel and Gupta, 1997). Anomalies of this type can be highly important when they cause explanations to be found and addressed. At the same time, it may emerge that the research was simply unreliable and that better metrics collected in other ways should be used. Not only must the accuracy and timeliness of information be established but its relationship to the achievement of management objectives must be proven.
- (5) **Managerial time:** marketers make money by marketing, not by reporting performance or planning. The time given to metrics should be balanced by the amount of improvement likely to flow from analysis and, perhaps more importantly, the better direction given to performance by internal alignment and marketing. Greater detail and insight must be balanced against the loss of simplicity and ease of access that it frequently implies.
- (6) **Credibility and compatibility:** one reason why marketers' efforts to evaluate and maximize their performance have not received more plaudits from the C-suite is that marketers have, to a large extent, been internally focused, on what they believe will lead to superior performance, not what is important to the consumers of this marketing metrics evaluation. Marketers need a common language with financial areas and senior management, one that has high standards of believability to them.

(7) Efficiency (achieving goals with the least use of resources) needs to be distinguished from effectiveness (whether the goals will be achieved at all).

These considerations provide criteria against which to test the value of different metrics in guiding marketing activity. In practical terms, management is advised to draw up a long list of potential metrics from the firm's own experience, that of competitors and peer companies, and the academic and practitioner literatures. Then each metric can be tested against these criteria to reduce the long list to a shortlist – i.e. the provisional dashboard. If the resulting number of metrics is too many metrics for regular review – e.g. more than twenty – then they can be divided into two groups: those that always appear and those that appear only when they fall outside preset controls.

When stripped to its barest level, shareholder value analysis suggests that the duty of the firm is to maximize the expected net present value of future income streams. One difficulty with this seemingly innocuous objective lies in the treatment of uncertainty. When planning the future, how should the firm value a defensive move that protects against a competitive move that may never eventuate? When evaluating the past, how does the firm distinguish between a bad decision and a bad outcome?

The evaluation of past performance exists primarily to drive the learning process, or feedback loop, in what Quinn, Anderson and Finkelstein (1996) call a “learning organization”. That in turn allows the firm to determine its optimal future strategy and marketing tactics.

Limitations and future research

We have not attempted to provide a methodology for selecting metrics for a given marketing organization. Rather, we have attempted to discuss popular but flawed metrics and considerations for the selection of metrics. The discussion has been at a theoretical level. Research is needed to establish which metrics typically best suit dashboards for which types of firm in which contexts.

Similarly, we have not explored the practical application of dashboards, such as, for example, which metrics should always appear and which only when they fall outside the prescribed range. The selection of benchmarks for comparisons is also part of the diagnostic process – e.g. comparison with prior years, norms (levels expected in plans) and competitors or the market as a whole. For the purposes of this chapter, we have simply taken the

dashboard to be the visible (to the CEO) top part of the measurement iceberg or warehouse.

Conclusions

This chapter has addressed the selection of metrics for senior management and the board. The diversity of objectives alone indicates the need for multiple metrics. Searching for a single “silver metric” is theoretically unsound, regardless of the fact that the candidates for such a role are in any case flawed.

The firm’s dashboard should bring together the appropriate measures, which will vary from organization to organization depending on its goals, its business model, the type of firm it is and the context within which it operates. The desire for comparability and an ability to establish benchmarks for performance drive the manager to a set of metrics shared with other organizations. The idiosyncratic nature of each organization’s environment, both internal and external, demands customized measures, however. Again, a compromise may well provide the solution. The key metrics that the organization uses to compare its performance may be drawn from best practice, industry norms and the firm’s own experience.

Our main conclusion is the need to separate performance evaluation from the comparison of future alternative marketing campaigns (planning). The two are divided by a learning phase in which lessons from performance are drawn before planning is undertaken. For continuity and consistency, the same metrics should be used when possible. We have explored the role of DCF metrics, such as customer lifetime value, customer equity and brand valuation. DCF techniques can be valuable for selecting strategy and planning and also for learning from experience – i.e. developing the business model. On the other hand, the balance of the argument seems to be that DCF techniques should not be used for performance evaluation. Return on investment promotes underperformance and short-termism, and should not be used at all. Return on customer is heavily dependent on forecasts; it appears simply to compare the cash flow for the year completed with the forecast for that year, and the consistency of last and this year’s forecasts for future years.

Our view is that, for managers to appraise and plan their businesses with insight and focus, they must go below these single summary measures to the few key drivers of the business, which may be moving interdependently but nevertheless provide a full view of the state of the business.

Appendix: Analysis of the ROC formula

The Peppers and Rogers definition of ROC can be expressed in a number of ways, but, after trying other variations that came to unlikely conclusions, this formula is probably what they intend:

$$\text{ROC}_t(\tau) = \frac{C_t(\tau) + \text{CE}_t(\tau) - \text{CE}_{t-1}(\tau - 1)}{\text{CE}_{t-1}(\tau - 1)}$$

where $C_t(\tau)$ is cash flow during the period ending at time τ as estimated at time t , and $\text{CE}_t(\tau)$ is customer equity at time τ as estimated at time t . Similarly, customer equity at time $\tau - 1$ ($\text{CE}_{t-1}(\tau - 1)$) is the customer equity (i.e. the net present value of future customer cash flows) at time $\tau - 1$ (that is, discounted to dollars at $\tau - 1$) as estimated at time $t - 1$. That is, $\text{CE}_{t-1}(\tau - 1) = C_{t-1}(\tau) + \text{CE}_{t-1}(\tau)$, i.e. the cash flow for period t plus customer equity with the calculation starting a year later. Residual values are assumed to be trivial and years are treated as units.

Substituting for $\text{CE}_{t-1}(\tau - 1)$ in the ROC formula above gives

$$\text{ROC}_t(\tau) = \frac{\{C_t(\tau) - C_{t-1}(\tau)\} + \{\text{CE}_t(\tau) - \text{CE}_{t-1}(\tau)\}}{\text{CE}_{t-1}(\tau - 1)}$$

In other words, this ROC formula does not measure return on the value of the marketing assets so much as the *excess* of the return for the current period (compared to forecast) plus any *increase* in forecast, both taken as a ratio of customer equity. This is directly analogous to the abnormal earnings growth used to value performance changes in other equities (see, for example, Penman, 2004, 201).

A difference is that this formula scales it by taking the ratio to incoming customer equity at time $\tau - 1$, $\text{CE}_{t-1}(\tau - 1)$. Of course, doing so introduces many of the problems of ROI, not the least of which is that maximizing ROC does not correspond to maximizing firm value.

References

- AMA (2005). *Marketing Accountability Study: White Paper*. Chicago: American Marketing Association.
- Ambler, T. (2003). *Marketing and the Bottom Line*, 2nd edn. London: Financial Times-Prentice Hall.

- Bahadir, S. C., and Tuli, K. R. (2002). *Measuring Marketing Productivity: Linking Marketing to Financial Returns*, Working Paper no. 02–119. Cambridge, MA: Marketing Science Institute.
- Bruno, H. A., Parthasarathi, U., and Singh, N. (2005). *Does Marketing Measure Up? Performance Metrics: Practices and Impacts*, Working Paper no. 05–301. Cambridge, MA: Marketing Science Institute.
- Collins, J. C., and Porras, J. I. (1994). *Built to Last*. London: Century.
- Debruyne, M., and Hubbard, K. (2000). *Marketing Metrics*, Working Paper no. 00–119. Cambridge, MA: Marketing Science Institute.
- Doyle, P. (2000). *Value-Based Marketing*. Chichester: John Wiley.
- Farrington, C. (2004). The language barrier. *Financial Times creative business*, 1 June, 15.
- Glazier, W., Nelson, R., and O'Sullivan, D. (2004). *Measures and Metrics: The CMO Council Report*. Palo Alto, CA: CMO Council.
- Gupta, S., and Lehmann, D. R. (2005). *Managing Customers as Investments*. Upper Saddle River, NJ: Wharton School Press.
- Kallapur, S., and Kwan, S. Y. S. (2004). The value relevance and reliability of brand assets recognized by U.K. firms. *Accounting Review*, 79(1), 151–72.
- Marketing Science Institute (2000). *2000–2002 Research Priorities: A Guide to MSI Research Programs and Procedures*. Cambridge, MA: Marketing Science Institute.
- McGovern, G. J., Court, D., Quelch, J., and Crawford, B. (2004). Bringing customers into the boardroom. *Harvard Business Review*, 82(11), 70.
- Mela, C. F., Gupta, S., and Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, 34(2), 248–61.
- Moorman, C., and Rust, R. T. (1999). The role of marketing. *Journal of Marketing*, 63 (Special Issue), 180–97.
- Morgan, N. A., Clark, B. H., and Gooner, R. (2002). Marketing productivity, marketing audits, and systems for marketing performance assessment: integrating multiple perspectives. *Journal of Business Research*, 55, 363–75.
- Morwitz, V. G., Steckel, J. H., and Gupta, A. (1997). *When do Purchase Intentions Predict Sales?*, Working Paper no. 97–112. Cambridge, MA: Marketing Science Institute.
- Olsen, S. (2005). Evaluating evaluation. Paper presented at Admap Monitoring Advertising Performance conference, London, 27 January.
- Peirson, G., and Ramsay, A. (1998). *Accounting*, 2nd edn. South Melbourne, Vic.: Marketing Science Institute.
- Peppers, D., and Rogers, M. (2005). *Return on Customer: Creating Maximum Value from your Scarcest Resource*. New York: Currency.
- Penman, S. H. (2004). *Financial Statement Analysis and Security Valuation*, 2nd edn. Boston: McGraw-Hill Irwin.
- Perrier, R. (ed.) (1997). *Brand Valuation*, 3rd edn. London: Premier.
- Quinn, J. B., Anderson, P., and Finkelstein, S. (1996). Managing professional intellect: making the most of the best. *Harvard Business Review*, 74(2), 71–81.
- Reibstein, D., Norton, D., Joshi, Y., and Farris, P. (2005). Marketing dashboards: a decision support system for assessing marketing productivity. Paper presented at the Marketing Science conference, Atlanta, 16–18 June.
- Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, 81(12), 46–55.

- Rust, R. T., Lemon, K. N., and Zeithaml, V. A. (2004). Return on marketing: using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1), 109–27.
- Schultz, D. E. (2000). Understanding and measuring brand equity. *Marketing Management*, 9(1), 8–9.
- Shaw, R., and Merrick, D. (2005). *Marketing Payback: Is Your Marketing Profitable?*. London: Financial Times Prentice Hall.
- Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118.
- Srivastava, R. K., Shervani, T. A., and Fahey, L. (1998). Market-based assets and shareholder value: a framework for analysis. *Journal of Marketing*, 62(1), 2–18.
- Venkatesan, R., and Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4), 106–25.

Andrew Likierman

Introduction

Qui ne risque rien n'a rien (Nothing ventured, nothing gained)

Companies are in the business of earning returns for shareholders as a result of taking risks, and we expect there to be a relationship between the two. Why put capital at significant risk for a return that is no higher than the return on government bonds? Or expect higher than average returns from low-risk activities? It is impossible to separate measuring the performance of a company from the risks that the management takes to achieve it.

Investors can reasonably expect greater rewards for specific risks, such as investing in start-ups or in parts of the world subject to civil war, than investing in more established and stable companies, in safe industries or in settled environments. Nonetheless, risk rarely figures when company performance is described. Financial measures are almost invariably given (“Our profits rose by 15 per cent”, “The company earned 23 per cent on invested capital”) without any risk context, even when comparisons are made or in any accompanying commentary. While average return is acceptable for an average risk, however, it is not for a highly risky venture. Without defining the context of risk, judgements about performance will be incomplete.

In most aspects of company operations, risk assessment plays a different, but equally important, role. It is an integral part of informed decision taking in achieving performance. Risk assessment is involved from the highest level in strategic choices about what activities to undertake, what assets to buy or what markets to serve all the way to detailed operational decisions about whether to accept payment in foreign currencies and the adequacy of safety measures in the workplace. It plays a part whether or not an organization is aware of managing risk, and many managers feel

that their instinct and judgement are enough – a behaviour risk. The danger is that this leaves company risk unplanned and unmanaged.

Managers are not always helped by those who form opinions. *Good to Great* (Collins, 2001), arguably the most influential management book of the first few years of the twenty-first century, does not even have the word “risk” in its index.

The evidence about what actually happens inside organizations is patchy. The financial services industry is, essentially, about managing different types of risk and has highly developed techniques for doing so, with regulators on hand to ensure that these are in place. Indeed, it is normal in large financial services companies to have a separate risk function. In part these mechanisms and the activities of regulators reflect past problems, including (in recent years) the bankruptcy of Barings Bank through derivatives losses, \$2.6 billion lost by a Sumitomo copper trader and \$0.7 billion lost by Allied Irish Bank in foreign exchange positions. It should be said that, while such losses are dramatic but infrequent, less publicized systemic failures in risk management can be as profound and far-reaching.

Other than in the financial services industry, the nature of the company’s activities and its size are probably the key determinants of whether there is systematic management of risk. At one end of the sophistication spectrum are probably the risk models of large energy multinationals, although large companies in many other industries are developing more advanced models following the requirement for listed companies in some countries to demonstrate effective controls, including control of risk.

The most stringent requirements on risk disclosure are those for listed companies in the United States by the Securities and Exchange Commission. In practice, many of the disclosures are anodyne or designed to cover every conceivable eventuality and thus avoid legal problems, but there is also a great deal of interesting information in the choice of risks disclosed, as well as the content of the statements. These are given in the box 12.1.

In the United Kingdom the control around risk is represented by the need for listed companies to explain if they don’t meet the “Turnbull” requirements as part of the Combined Code (FRC, 2003). More significantly, the Operating and Financial Review – currently best practice but at one stage a potential requirement for listed companies – requires “a description of the principal risks and uncertainties facing the company” (ASB, 2005, 26).

Box 12.1: The influence of SEC risk disclosure requirements

The requirements apply to non-US companies listed in the United States as well as to US companies. Two European examples follow.

- Nokia is listed in the United States and, following the format in its filing, has two pages in its annual report listing risks, including markets, technology, manufacturing, suppliers, products, employees, macroeconomic factors, financing, health and safety and regulation.
- Phillips has no fewer than four pages in its annual report covering risks, including its overall approach to risk, market factors (focusing on China), competitive factors, strategic alliances, technology, business partners, personnel, financing, product liability, supply chain, legal aspects and pension fund matters.

Many of the comments above are based on evidence from larger firms. By contrast, very little evidence is available about how risk is managed for small and medium-sized enterprises (SMEs), though there is highly visible circumstantial evidence of widespread failure to manage risk through a high attrition rate, particularly of start-ups.

Those that manage to survive have, by implication, been at least risk-aware, since it is very difficult to continue to exist as an SME over an extended period without such an awareness. There may be formal risk management only in such areas as checking the creditworthiness of new clients or ensuring that health and safety requirements are met. Otherwise, the approach to risk management is likely to be simply an aversion to loss, precisely because of the sense of vulnerability surrounding SMEs.

Why is risk management not more firmly on the management agenda for many companies, large or small? Some possible reasons might be as follows.

- Risk management has developed relatively recently, so managers are not aware of it.
- Organizations do not have incentives for managers to take account of risk on a systematic basis.
- It is too complex for an older generation of senior managers to understand.
- Those outside financial services regard it as being to do with complex financial instruments, with nothing to offer them.
- Both quantitative and qualitative techniques are dismissed as being of dubious value and inferior to the personal judgements, based on personal experience, that managers feel they bring to bear.
- Risk management is actually taking place, but is not formally recognized as a process.

All these reasons are worrying, because essential elements of risk management are awareness of and communication about the balance of risk. Even

those who feel that they can do better than what they see as flawed or redundant models need to be open-minded about what a more systematic approach might offer.

Those in larger companies may not have much choice. The formal requirements for listed companies, together with pressure from audit committees and the non-executive directors, are pushing the risk agenda into the organization. The professionalization of risk management is providing additional impetus to do so.

In making the links between performance and risk, it's important to remember that, while pressures and techniques are part of the universal world of business, approaches to dealing with the pressures and applying the techniques vary greatly between business cultures (Schneider and Barsoux, 2003). To give just one example, the extreme consequences of risk taking for performance – great wealth or bankruptcy – are more acceptable in the United States than in many European countries.

One of the differentiators here is the role of “face” (Earley and Ang, 2003, 39); generally, there is less risk appetite in countries where the social disgrace associated with failure is seen as a severe sanction. Triandis (2002) notes the greater importance of face in collectivist cultures (e.g. Japan and South Korea) compared to those that are individualist (e.g. Australia and the United States). Approaches also vary between industries. One would expect a different attitude in the mining industry from accountancy, and individual organizations vary from the gung-ho to the risk-phobic.

The sections below deal with these perspectives of links between risk and performance. Section 1 deals with risk in assessing overall company performance while section 2 sets out internal risk management to improve performance. In one aspect of company operations, however, performance and risk are linked rather differently. This is in the way performance is linked to individual reward through any performance-related scheme. Such schemes have to link incentives with amounts at risk for the individual. Section 3, therefore, discusses the links between risk and reward for individuals.

Risk in each case is taken as “any unintended or unexpected outcome of a decision or course of action” (Ansell and Wharton, 1992) or, more succinctly, “any deviation from expectations”.¹ Figure 12.1 illustrates the relationship between the three areas.

¹ Correspondence between author and Zeger Degraeve.

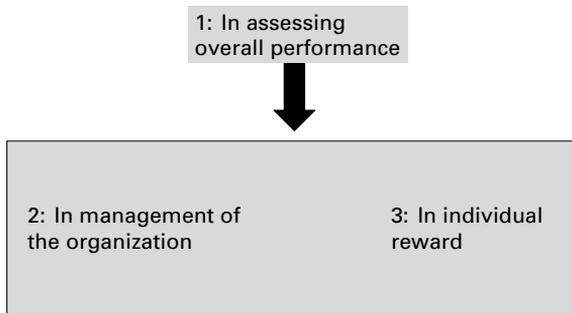


Figure 12.1: Risk and performance: three perspectives

Risk in assessing overall company performance

The directors recognise that creating shareholder value is the reward for taking and accepting risk. (Tesco Annual Report)²

One of the weaknesses of traditional performance measures, such as return on investment, whether profit-based (e.g. return on assets) or cash-related (e.g. “free” cash flow), is that they don’t take account of risk. Performance comparisons are all too often made as if the organizations take equal amounts of risk. However, this is obviously absurd. Baghdad is a riskier place to do business than Bangalore and high-yield junk bonds are riskier than US government stock.

It might be thought that at least comparisons with others in the same industry might be made without reference to risk. In other words, comparing return on capital employed for the oil majors, or profit growth for pharmaceutical companies, implicitly takes into account that those in an industry bear similar risk. In practice, though, there will rarely be a close comparison, since in most industries the mix of businesses varies greatly between large companies. Even when the industries are not diversified by type of business, there is greater or less exposure to certain kinds of risk, whether it be to sources of supply in the case of oil, or the type of medical problem on which research is focused in the case of pharmaceuticals.

Comparing profit or cash flow with companies in dissimilar industries or at different stages of development is of particularly doubtful value in

² All references to Annual Reports are to 2005 issues, covering 2004. In each case, a full and updated report is available on the company website.

measuring company performance, unless relative risk is taken into account. To compensate for higher risk, in general a start-up has to do better than an established player and a biotechnology company has to earn higher returns than a water utility.

One way to take account of risk is to incorporate it in an appropriate performance measure, such as through calculating the risk-adjusted return on capital. The use of one of the economic profit models that compares returns with the cost of capital will enable risk to be reflected in the cost of capital used. In such models, accounting profits are adjusted to give financial returns that will provide the appropriate signals for current decision taking and for measuring long-term cash returns. These returns are then compared to the weighted average cost of capital, the risk factor being incorporated in the cost of equity (see Martin and Petty, 2000, for different types of adjustment and Arnold, 2005, 93–105, or Ogier, Rugman and Spicer, 2004, for a full description of the methodology). Box 12.2 gives an example.

Box 12.2: Economic profit at Michelin

Michelin provides not only an economic profit figure each year but a comparison with the target. In the five years to 2004 the actual return varied from 6.5 per cent to 8.5 per cent. The target varied from 9.8 per cent to 11.4 per cent – a rare example of detailed disclosure. The commentary gives a division of the capital employed into “economic capital” and debt. It also gives the cost, which in this period varied around 16 per cent for economic capital, while the cost of debt fell from around 7 per cent in 2000 to less than 5 per cent in 2004. The notes to the table in the Annual Report explain the basis of these figures.

Source: Annual Report.

There are different economic profit models, including economic value added, cash value added, cash flow return on investment, free cash flow and others. These are marketed by specialist consultants, each with a different approach to making adjustments to the profit figure to get a return that is nearer to cash and less affected by accounting policies.

The results, by whatever method is chosen, can then link to decision taking inside the organization, as set out in the section below. These results should make clear the methodology and be accompanied by a commentary to ensure that the assumptions are clear and the numbers are put in perspective.

For investors in quoted companies, risk can be taken into account by comparing total shareholder return (TSR) – capital growth plus dividends – to the risk-adjusted rate of return, using the company’s β , which indicates,

based on past performance, the sensitivity of a company's return to the return on the market portfolio. The calculation is to add the risk-free rate of return (based on government bonds) to the equity risk premium (the expected return in excess of the risk-free rate), multiplying the latter by the company's β .

To illustrate what this means in practice, at the time of writing the approximate β for Cadbury Schweppes was 0.5, the value for Whitbread 1.0 and that for Reuters 1.5 (*Risk Measurement Service, 2005*). These figures indicated that Whitbread's shares tend to move in line with the market, Cadbury Schweppes' move much less and Reuters' much more. In investment terms, Cadbury Schweppes would be regarded as a defensive share, Reuters a relatively risky one. Taking assumptions of a risk-free rate of 5 per cent and an equity premium of 5 per cent, we would compare the TSR of Cadbury Schweppes to 7.5 per cent, Whitbread's to 10 per cent and Reuters' to 12.5 per cent. In practice, there are further complexities attached to the calculation (set out in corporate finance texts, such as Brealey, Myers and Marcus, 2004), but these do not affect the basic principle behind the calculation.

The Sharpe ratio, which compares return with the volatility of return, is also commonly used for financial portfolio evaluation, though it can be applied more widely. It is used to allow the comparison of a risk-adjusted return with different types of assets over time. The higher the ratio the better the return in relation to the level of risk relative to volatility (see box 12.3 for an example).

Box 12.3: The Sharpe ratio

The Swedish holding company Investor AB publishes its Sharpe ratio data. Over a ten-year period to the end of 2004 it was 0.26. For 2004 itself the ratio was 1.2, much higher, and therefore more favourable.

Source: Annual Report.

The measures above provide investors with the means to construct portfolios of quoted companies based on a desired balance of risk using mathematical relationships. For quoted companies, information related to the stock market may also be of use to investors in assessing risk, although there is also a great deal of "noise" from other factors that affect share prices, meaning that the information can only be indicative. For example, a dividend yield that is well above the sector average may indicate greater risk, because the company's share price is vulnerable or that the dividend is at risk. Another is the forward price/earnings ratio.

A ratio that is very high or very low compared to the sector implicitly reflects an assessment of the relationship between risk and reward. A third possible signal is the balance of analyst buy/hold/sell recommendations and the accompanying commentary. To illustrate what these measures mean for real companies, at the time of writing BP had coverage of thirty-three analysts and a buy/hold/sell balance of 15/15/3, while Shell had coverage of twenty analysts and a balance of 5/10/5. (Note that the balance is not always tilted towards buying: Invensys, an engineering company, had a balance of 1/5/6.)

As far as P/E ratios are concerned, the average of shares traded on the New York Stock Exchange was around 20. Google, however, was trading at a ratio of 75. Holders of Google were, at this price, implicitly taking an equally clear view of risk and reward in buying seventy-five years of last year's earnings compared to those buying twenty years of the market average. (Evidently, the analysts thought so, since the buy/hold/sell balance was 25/8/1.) So, indeed, were holders of Biadu.com, the "Chinese Google", which at one time traded at a P/E ratio of more than 2,000. These ratios (including Baidu.com) are neither cheap nor expensive. Rather, they implicitly reflect a different set of assumptions about risk and future reward.

Dividend yields and forward P/E ratios relative to the sector and the buy/hold/sell balance can only be indications, being affected by so many other factors. Thus, BP's better rating by analysts, manifest both in comments and a higher P/E ratio, for example, is an implicit indication of confidence in BP's ability to handle risks relative to Shell, other things being equal. However, it may also be due to many other factors, certainly in the short term.

The degree of imprecision will, of course, be greater for unquoted companies, since the absence of a market price makes comparisons more complex and takes away the means to provide a company's β (Feldman, 2005, 69–89). Risk will therefore be assessed more subjectively – one of the reasons for the discount attached to unquoted companies. Even a notional P/E ratio based on comparisons to the quoted sector can be interpreted only in light of an assessment of risk.

By contrast, lenders have very well-established measures on default risk based on models using historical trends. The major rating agencies, as well as financial institutions, provide continuous and closely watched risk criteria. Standard and Poor's, for example, provides carefully graded opinions for different types of obligation. The three top grades (AAA, AA, BBB)

range from “the highest” to “adequate” and the four low ones (BB, CCC, CC, C) from “ongoing uncertainties” to “highly vulnerable”; the lowest (D) is already in default. Each rating may be qualified by a “+” or “-”. Ratings are given for different countries and evaluate outlook as well as the current position.³

The risk of insolvency is covered by a variety of models. One that has been available since the 1960s is the “Z score”, which brings together sets of financial variables to give a mathematical probability of default. There are many others (see Edwards, 2001).

The variety of techniques outlined in this section suggests that risk assessment should be part of every assessment of company performance. Limitations of the techniques indicate that they should be used with caution, not that they should be ignored.

Internal risk management and performance

There are areas of the group’s business where it is necessary to take risks to achieve a satisfactory return for shareholders, such as investment in R and D and in acquiring new products or businesses. In these cases it is the group’s objective to apply its expertise in the prudent management rather than elimination of risk. (GlaxoSmithKline Annual Report)

What are the areas in which risk management is relevant to performance? They include the following.

- (1) Identifying vulnerabilities and opportunities in strategy and budgeting, including capital budgeting, in light of risk appetite, using such techniques as Monte Carlo simulation to create a risk profile.
- (2) Assessing the possibilities of cost and time overruns in managing projects, again bearing in mind risk appetite, using both formal probability and β analysis as well as more subjective methods (Arnold, 2005, 11–51).
- (3) Incorporating risk in target setting and performance measures; for example, incorporating key risk indicators as part of a balanced scorecard (say, missed training days in an insurance company as a leading indicator of potential misselling).

³ See “Standard and Poor’s rating definitions” at standardandpoors.com.

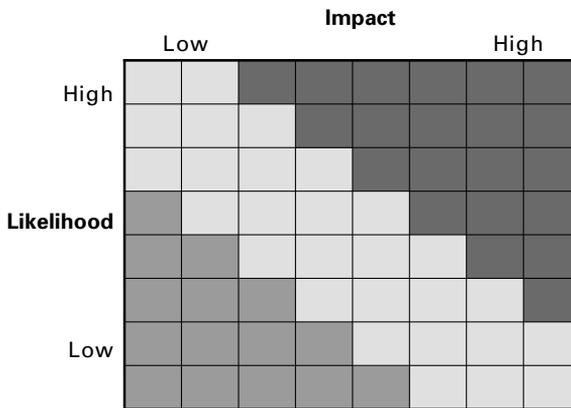


Figure 12.2: Risk heat chart

- (4) Systematically managing operational risk at all levels through good management practices as well as special techniques such as the use of real options (Risk Books, 2003) to adjust operating decisions in response to uncertainty.
- (5) Varying the risk factor in relation to specific projects in the WACC used, rather than using a single WACC for all projects.
- (6) Identifying risk probability (the likelihood that a particular outcome will occur) and impact (the effect a particular outcome will have if it does occur) in continuing operations, for example using “heat charts” to show the relationship between the two. Figure 12.2 gives an illustration.

Managing these risks is a core differentiator in the quality of management. Bad managers are bad because they allow a company to accept risks without being aware that they are doing so. Skill here includes a systematic approach, communicating that approach and a willingness to review risk appetite in the light of changing circumstances.

As already noted, risk management has traditionally been seen as the province of the financial sector and some very large companies in sectors with high levels of perceived risk. It is usually framed in terms of financial instruments, and, indeed, the spectacular growth of parts of investment banking has been based on developing ever more sophisticated financial instruments to deal with risk.

Since their operations are all about managing the risk of their borrowing and lending portfolios, it is certainly the case that the performance of companies in the financial sector is closely linked to their success in risk

management. Even before being reinforced in recent years by external requirements, particularly from the Basel Committee on Banking Supervision, there were more established techniques, including the use of value at risk, or VaR (see box 12.4), than for other kinds of companies.

Box 12.4: Value at risk

- A key measure of risk, especially in financial service companies.
- “The maximum loss over a target horizon such that there is a low, prespecified probability that the actual loss will be larger” (Jorion, 2005, 253).
- Normally expressed as a monetary amount, given a confidence level and a period of time. For example, the average total daily VaR at Barclays Capital in 2004 (Annual Report, 62–3) was £34.3 million at a confidence level of 98 per cent. Daily losses over this level are therefore likely to occur, on average, twice in every 100 days. This total was the result of aggregating daily VaR for five types of risk.
- Basel rules require a specific calculation of risk (over ten trading days at 99 per cent confidence levels) for all banks, with additional safeguards against failure.
- Caution is needed in using and interpreting the figures.
- VaR does not cover worst loss, or the distribution of losses in the “left tail”, and stress testing is therefore required.
- VaR is not a substitute for overall good risk management (Culp, Miller and Neves, 1999).
- Variations of the same principle are included in such measures as “cash flow at risk” and “earnings at risk”.

By contrast, the techniques for managing risk in other sectors are far less developed, for many of the reasons set out in the previous section. Managers may claim to be taking account of risks anyway. Their confidence will be misplaced, however, if they recognize only what’s most easily identifiable and quantifiable, and their decisions will be skewed unless they balance the downside with the upside. Contrast the sustained success of the US company General Electric in managing risk through profitable growth in the latter part of the twentieth century with its British namesake (later Marconi) over the same period. The latter was first weakened by twenty years of failing to take enough risk then almost brought down by excessive financial and strategic risk taking in the four years to 2001.

Risk affects all aspects of a company's operations. Companies are most aware of the implications in the price at which they can raise finance. However, it also affects their reputation ("Are they trustworthy?"), ability to sell ("Will they still be around if anything goes wrong with the product?"), ability to buy ("Will they pay?") and ability to hire and pay staff ("How secure is my job?").

The board's skill in identifying what risks they are prepared to accept to achieve the company's objectives – risk appetite – and then in managing that risk appetite is fundamental to a company's strategy. Contrast the implicit risk appetite behind two sets of management thinkers: the "big hairy audacious goals" of Collins and Porras (1996) and the "fast second" approach of Markides and Geroski (2005), who suggest that rewards can be reaped by allowing others to take the risks first.

Companies often associate risk management with avoiding problems, particularly running out of cash. What's relevant to performance, though, is the skill in managing the risk, not avoiding it. As the example of GEC in the United Kingdom showed, there can be serious consequences from taking an over-conservative attitude to risk, which can give competitors as great an advantage as if a reckless attitude is taken. These consequences come about from circumstances as varied as failing to grasp strategic opportunities, failing to hire good enough people or keeping excess capital.

"Managing" here means choosing between eliminating, tolerating, minimizing, diversifying, hedging, transferring or insuring risks (Degraeve, 2004, 48). Many techniques, such as insurance, are very familiar (and very long-standing; as Kay has pointed out (2003, 232–3). *The Merchant of Venice* would have been a very different play if marine insurance had been available to Antonio). Others, such as hedging, become familiar at times of significant swings in currencies or raw material prices. Some, such as ways of transferring risk, are sometimes less familiar, since they can appear in complex technical products, such as credit risk derivatives.

The areas to be managed range from managing foreign exchange exposure to the use of risk-adjusted discount rates for investment. They include the management of long-term contracts, which could expose the firm to significant losses, whether as supplier or customer. They extend to choosing a pension scheme, when there is a risk to the company of having to make good a deficit.

An illustration of how wide-ranging the definition of the areas to be covered and the actions to cover them can be is given in box 12.5 for

Land Securities plc, a major UK property company. Note that the word “mitigation” rather than “management” is used.

Box 12.5: Risk mitigation at Land Securities

The company identifies twelve types of risk and the ways in which it seeks to deal with them. Among the “mitigation activities” are:

- internal and external research (for market risk);
- key performance measures (for asset management risk);
- compliance procedures and participation in industry organizations (for regulation risk);
- monitoring and the use of financial limits (for development risk);
- an active HR policy (for human resources risk);
- back-up plans (for IT risk); and
- relationship management programmes (for stakeholder relationship risks).

Source: Annual Report.

In identifying risks to be managed, the conventional categorization into strategic, financial and operational is a useful starting point, but it is wise not to get hung up on classification. For a start, the scope and definition of terms differs; legal risk is sometimes identified separately, sometimes as part of operational or financial risk. A second reason is that risks may be related to each other, as when unhedged raw material price fluctuations affect operations. Finally, risks may change, as when a collection of small operational risks becomes strategic.

The key to success of risk management as a whole – for medium-size and smaller enterprises as much as for international companies – is to embed it into day-to-day operations and line management. For example, overall company targets must be set as part of the corporate strategy and in annual budgeting to reflect risk appetite.

As already noted above, one of the ways in which companies can embed risk in decision taking is through the calculation of economic profit. Box 12.6 has an example of the principles, though with some simplifying assumptions. By incorporating the cost of capital adjusted for risk, techniques such as EVA[®] provide a link to shareholder value as well as a clear basis for choosing between alternatives. Nonetheless, the use of such techniques remains limited, and many companies prefer to rely on less formal methods.

Box 12.6: Economic profit

- (1) Economic profit = adjusted post-tax earnings less capital x weighted average cost of capital; here take \$50 post-tax earnings – \$200 capital x 8 per cent WACC (see 2) = \$34.
- (2) WACC includes a cost of debt and equity; here take \$200 capital as being 50 per cent debt @ 6 per cent cost of debt net of tax relief and 50 per cent equity @ 10 per cent cost of equity (see 3) = 8 per cent WACC.
- (3) Cost of equity derived from risk-free rate + market risk premium $\times \beta$; here take risk-free rate as 5.5 per cent, market risk premium as 3 per cent and β as 1.5 = 10 per cent.

Improvement through greater formality does not necessarily mean more numbers. The assessment will be best through a combination of numbers and commentary to put them in context. Indeed, there is a particular danger of relying on a single risk measure, since each has limitations, as shown with value at risk, in the same way that profit or cash flow cannot be relied on as single performance measures.

How can a judgement be made about how well an organization has managed risk? Making sure that processes are in place – risk reporting on which action is taken, risk awareness surveys giving satisfactory results compared to others, etc. – is a starting point.

While the right processes are essential, the key measure of success has to be the outcome achieved by those processes. This does not mean only the frequency and severity of losses, since if the frequency and severity of losses are less than anticipated it could be that the opportunities lost have been omitted and the cost of prevention could have been excessive.

The more balanced way to make the assessment is to look both at what has gone wrong and what has gone right to see whether the opportunities taken and the frequency and severity of any losses or problems correspond to the risk appetite identified. If the frequency and impact of opportunities taken and losses suffered are much greater or less than planned, it provides at least some evidence that risk has not been well managed. Realistically, this will be more of a qualitative than a numerical assessment, not least because of the difficulties of fixing a baseline number and of choosing the time period over which to make the assessment.

In making the assessment, risk culture – the success of the company in communicating risk appetite, linking it to incentives and ensuring that it is

understood and used – needs to be taken into account. Having techniques in place is a necessary condition to demonstrate that risk is being managed but not a sufficient condition for risk management to be successful, since it depends on how the techniques are applied.

Risk and individual reward

Incentives should align the interests of executives and shareholders and reward the creation of long-term value within a framework which enables risk to be assessed and managed (one of ICI's principles for remunerating executive directors). (ICI Annual Report)

The third area in which risk and performance are linked is in relation to performance-related reward. The risk element applies both to companies and to the individuals who work in them.

For companies, the risks are an element in well-established issues of principal–agent relationships. These are “general to virtually all cooperative activity among individuals” (Jensen, 2000, 137), the shareholders as principals taking a risk in the use of employees as agents. Incentives within employment contracts are designed to help align interests, including the different appetite for risk of shareholders and employees.

For individuals, the incentive element designed to encourage them to deliver better performance for the shareholders puts their future income at risk. With an implicit assumption that they are risk-averse, employees need to be compensated by a risk premium for the uncertainty attached to their future income. Individuals' risk is exacerbated if there are “inefficient” contracts – those that do not balance the cost of risk bearing against incentive gains. The performance element in contracts may also be affected by difficulties in defining outcomes and disparities in information about them, giving rise to two phenomena much discussed by economists, for example in Milgrom and Roberts (1992): adverse selection (taking advantage of information disparities) and moral hazard (when actions are not freely observable).

Many schemes seek to align the risks and rewards of employees and shareholders, of which issuing share options to senior executives is the most prominent. In significant respects, however, the relationship between risk and reward for the two groups is very different. Shareholders bear far less risk (Norris, 2005). Not only can they diversify their portfolios but they can also sell their shares, though only with constraints if the shares are not

publicly quoted. Employees will tend to be less tolerant of risk, being unable to diversify their own employment risk in the same way as shareholders and being vulnerable to individual projects going wrong (Rappaport, 1998). Leaving employment is also a far bigger personal risk than when a shareholder sells shares. Incentive schemes covering groups carry their own risks, notably a further distance between individual effort and reward, including the danger of carrying “free riders”.

Most public attention in this field is focused on executive reward in large, quoted companies. High headline payouts, especially when profits fall, inevitably give rise to accusations of payments for failure, with the message that risk and reward have not been linked – as in the stinging attack on current practices by Bebchuk and Fried (2004).

Managing the risk–reward relationship and providing clarity about incentives and performance is at the heart of remuneration committees’ (RemCos’) role. They must be seen to address these issues in setting up schemes – in economic terms, setting an efficient contract balancing risks against incentive gains. However, there are plenty of banana skins for RemCos (Likierman, 2006).

One is failing to identify the risk conditions. This becomes a significant issue when, as in the late 1990s, huge share price gains give rise to executive payouts linked only to apparent market over-exuberance. The areas of dissatisfaction have included targets being set too low, the incentive amounts being set too high and the link to performance being too tenuous.

Another is approving schemes in which the downside for the individual relative to the risk taken is less than the upside. This issue has attracted a great deal of attention, particularly in relation to share options in the United States and the United Kingdom. An example is offering options simply to align shareholders’ interests with those of senior executives rather than as any kind of incentive. Another is repricing options after share prices fall, giving executives an apparently one-way bet. The use of share options has decreased following changes to their tax and accounting treatment, but the problem can exist with other forms of incentive scheme (Rich, 1999).

These issues are long-standing, but they have become far more prominent as the focus on good corporate governance over the last fifteen years has increased disclosure, and therefore the potential for reputational damage to the company. The problem of reputation is much less for companies that are not publicly quoted, but linking individual reward and performance in such companies is often no less contentious internally. Hiring remuneration consultants has therefore been seen as a way not only to improve schemes

but also to limit external reputational risk and internal discontent by being seen to be in line with others.

Other issues, such as the way targets are set, apply at all levels. An ability to balance risks and rewards in a way seen to be fair is the basis of any good incentive scheme, and they have become more sophisticated in response to pressure (see, for example, Chingos, 2002, for a guide to the factors involved). All the same, widespread and continuing doubts about such schemes and the assumptions that underpin them (see, for example, chapter 21) show that this is a difficult area for the HR departments of companies to get right.

Conclusions

In the book *What Management Is*, Magretta (2002, 173) comments that “without innovation and risk taking, there would be no economic progress. The discipline of management helps to increase the odds that the risky business of innovation will pay off.” It is effectively impossible to disentangle risk from management, or to argue that risk should not be taken into account in performance assessment and in decision taking.

This chapter has shown three perspectives of the relationship between risk and performance.

- (1) Company performance comparisons must incorporate a risk element, if possible in the measure used and in any case in the commentary.
- (2) Risk management must be integrated into the business, including performance measurement, not treated as an optional extra. Risk management does not mean avoiding risks but it does mean accepting that not all risks can be quantified.
- (3) Risk must be treated as an essential element in performance-related reward. Some investors and many more managers believe that systematic risk assessment is superfluous. They believe that experience and instinct will see them through. It's a high-risk assumption.

References

- Ansell, J., and Wharton, F. (1992). *Risk: Analysis, Assessment and Management*. Chichester: John Wiley.
- Arnold, G. (2005). *Handbook of Corporate Finance*. London: Financial Times Prentice Hall.
- ASB (2005). *A6 of Operating and Financial Review*. London: Accounting Standards Board.
- Bebchuk, I., and Fried, J. (2004). *Pay without Performance*. Boston: Harvard University Press.

- Brealey, A., Myers, S. C., and Marcus, A. J. (2004). *Fundamentals of Corporate Finance*. Boston: Irwin.
- Chingos, P. T. (2002). *Paying for Performance*. Chichester: John Wiley.
- Collins, J. (2001). *Good to Great*. New York: Random House.
- Collins, J., and Porras, J. I. (1996). Building your company's vision. *Harvard Business Review*, 74(5), 65–77.
- Culp, C. L., Miller, M. H., and Neves, M. P. (1999). Value at risk: uses and abuses. In Brown, G. W., and Chew, D. H. (eds.), *Corporate Risk: Strategies and Management*, 239–48. London: Risk Books.
- Degraeve, Z. (2004). *Risk: How to Make Decisions in an Uncertain World*. Normich: Format.
- Earley, P. C., and Ang, S. (2003). *Cultural Intelligence: Individual Interactions across Cultures*. Stanford, CA: Stanford University Press.
- Edwards, B. (ed.) (2001). *Credit Management Handbook*. Aldershot: Gower.
- Feldman, S. J. (2005). *Principles of Private Firm Valuation*. Chichester: John Wiley.
- FRC (2003). *The Combined Code on Corporate Governance*. London: Financial Reporting Council.
- Jensen, M. C. (2000). *A Theory of the Firm*. Cambridge, MA: Harvard University Press.
- Jorion, P. (2005). *Financial Risk Manager Handbook*. Chichester: John Wiley.
- Kay, J. (2003). *The Truth about Markets*. Harmondsworth: Penguin.
- Likierman, A. (2006). Measuring the success of the remuneration committee. *Benefits and Compensation International*, 35(6), 3–6.
- Magretta, J. (2002). *What Management Is*. London: Profile.
- Markides, C., and Geroski, P. A. (2005). *Fast Second*. San Francisco: Jossey-Bass.
- Martin, J. D., and Petty, J. W. (2000). *Value-based Management*. Boston: Harvard Business School Press.
- Milgrom, P., and Roberts, J. (1992). *Economics, Organisation and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Norris, P. (2005). Shareholders' attitudes to directors' pay. In Tyson, S., and Bournois, F. (eds.), *Top Pay and Performance*, 29–56. Burlington, MA: Elsevier.
- Ogier, T., Rugman, J., and Spicer, L. (2004). *The Real Cost of Capital*. London: Financial Times-Prentice Hall.
- Rappaport, A. (1998). *Creating Shareholder Value*, 2nd edn. Boston: Free Press.
- Rich, J. T. (1999). Aligning executive pay and the creation of economic value. In Risher, H. (ed.), *Aligning Pay and Results*, 80–100. New York: American Management Association.
- Risk Measurement Service (2005). *Risk Measurement Service*, 27(2).
- Risk Books (2003). *Advances in Operational Risk: Firm-wide Issues for Financial Institutions*, 2nd edn. London: Risk Books.
- Schneider, S. C., and Barsoux, J. L. (2003). *Managing across Cultures*. London: Financial Times-Prentice Hall.
- Triandis, H. C. (2002). Generic individualism and collectivism. In Gannon, M. J., and Newman, K. L. (eds.), *Blackwell Handbook of Cross-cultural Management*, 16–45. Oxford: Blackwell Business Books.

Rob Austin and Pat Larkey

Introduction

The explosive growth of the information-related and science-based sectors of industrialized economies has seen concomitant growth in the demand for products and services with potential to help organizations apply what they know more profitably. There are entirely new occupational categories – combinatorial chemist, network manager, software engineer, to name just a few – that produce economic value mostly by creating and manipulating symbols. Even in traditional occupations, ways of working have come to depend on intellectual activity more than physical activity. Factory workers invent and share process improvement ideas; salespeople develop novel ways to use the Web to reach or retain customers.¹

As the nature of work has changed, the relationship between an organization's measured resources and its market success has become tenuous. Factors that contribute substantially to a firm's market success largely elude traditional means of quantification. Unlike materials and equipment, core competencies comprised of the distinctive abilities of employees and teams are not easily captured on balance sheets. This fact not only makes valuation of individual firms harder, it also complicates traditional analyses for allocating resources, improving processes, and compensating employees.

Existing notions of performance measurement and organizational control meet with substantial challenges in knowledge work settings. In this chapter, we identify distinctive characteristics of "knowledge work"² that impact our

¹ The ability to create, retain, communicate and use knowledge is often critical to organizational success (Davenport, 2005; Davenport and Prusak, 1998; Nonaka and Takeuchi, 1995). Management of knowledge has become a frequently acknowledged source of "core competencies" (Prahalad and Hamel, 1990), which themselves give rise to sustainable competitive advantages (Leonard, 1995).

² "Knowledge work" is, in many ways, an unsatisfactory phrase, because it has become vague in meaning from overuse. We use the term simply because it is widely recognized and conveys some sense of what

ability to measure it, and briefly discuss the research challenges and practical implications associated with those characteristics. Conceptually, these characteristics can be placed in three categories.

First, knowledge work is less *observable* than physical work. The activities involved in loading coal into a railway wagon are easier to see, understand and evaluate than the activities involved in progressing towards a successful database design. Although the problem of *observability* has long been known and studied in measurement settings, it takes on new dimensions when the activity being measured is intangible “thought-stuff”.

Second, the *motivation* of knowledge workers is, arguably, more reliably intrinsic than it is for many physical workers. Knowledge work is often idiosyncratic and oriented towards problem solving, hence intrinsically interesting for workers; physical work is more often repetitive and oriented towards compliance, hence less interesting.³ The possibility that strong intrinsic motivation can be relied upon to direct workers’ activities requires rethinking research based on less optimistic assumptions about human motivation (such as economic agency theory). The social psychology literature on human creativity underscores the importance of this possibility by describing a complex relationship between motivation and the creativity of work outputs (e.g., Amabile, 1996).

Third, a high degree of individual capability in the worker is critical to knowledge work success, whereas for the physical worker, success more often depends on his or her consistent compliance with a plan or efficient participation in a system.⁴ An excellent technical support person is excellent not because of his or her performance on well-known and well-understood tasks but, rather, because he or she is good at the exploration, knowledge creation and analysis needed to perform excellently on unprecedented problem solving

it means. For our purposes, knowledge work is work in which important value-creating transformations occur in the realm of ideas or symbols; or, alternatively, in which a substantial amount of productive activity is intellectual rather than physical. We assume that this work will tend towards conceptual complexity in ways that create disparities among work participants in the understanding the work itself, and the events that occur in the course of completing the work.

³ This is, of course, not always the case. An excellent carpenter no doubt does physical work that very much interests her/him, and his interest no doubt plays a vital role in developing his skills to a high level. However, such an excellent carpenter may well have introduced a substantial element of knowledge work into his efforts, in the form of, say, creative improvements in methods or cabinet design. DeMarco has pointed out that the technologies we have available for automating routine or repetitive symbolic tasks are very good, much better than the technology for accomplishing the same kinds of physical work (Austin, 1996); therefore we quickly tend to automate uninteresting knowledge work.

⁴ This is not true for all physical work, obviously. Professional athletes do work that is very physical and requires very highly developed individual capabilities in the “worker”.

or other vaguely defined tasks. In knowledge work, *talent*, *skill* and *knowledge differentials* (which we abbreviate as “TASK”) matter a lot; performance is much more about what a worker can do than about what you can get him or her to do. For knowledge work, identifying appropriately skilled individuals and inducing them to join the organization are often at least as important as managing them once they are in place. Once in place, managing these employees in a way that magnifies and leverages their skills rather than mutes them, and allocates resources to them appropriately, is also a substantial challenge.

There are sizeable literatures in fields as diverse as accounting, economics, industrial engineering, organization theory, psychology and sociology that are useful in thinking about observability, motivation and TASK differentials. Because of the sheer volume of this research, it is not possible to cover these subjects exhaustively. Rather, in this chapter we draw selectively from various fields the findings that bear most directly on the problems of observability, motivation and TASK differentials in measuring knowledge work.

The problem of observability

Observability problems – that is, problems in actually discerning the essence of performance in a phenomenon we are attempting to measure – have been much studied in physical and generic organizational contexts.

Economic agency theories (Ross, 1973; Jensen and Meckling, 1976; Holmstrom, 1979; Holmstrom and Milgrom, 1987, 1991; Baker, 1992) constitute a relatively recent and direct attempt to address issues of observability, which are notable for their growing prominence in research and practical contexts (they are, for example, increasingly used in discussions of how executives should be paid for their performance). The foundational model in this category was introduced by Ross (1973) and later refined by Holmstrom (1979). Essentially, all subsequent theoretical models of this type are elaborations of the “Ross–Holmstrom” (R–H) model. The model attempts to distil observability problems into their simplest, most fundamental elements.

The R–H model depicts an organization in drastically simplified form: two individuals and the contractual relationship between them. A *principal* has control over productive resources and seeks to hire an *agent* to do work that will transform resources into output with money value. The principal wants to maximize “profit” – the difference between the monetary value of the agent’s output and the payment required to induce effort from the agent. The agent’s

motivations are equally self-interested and almost as simple: he wants to maximize his income while minimizing effort and risk.⁵ The interests of the two are therefore opposed. The principal wants to extract as much value as possible from the agent and yet pay him or her as little as possible. The agent wants to do as little work and bear as little risk as possible, and yet be paid as much as possible.

Observability problems enter into the model through an assumption that the principal cannot directly observe the agent as he/she works. The justifications of this assumption are as follows:

- (1) Often it is not feasible for a manager to watch an employee at every moment, and in any event doing so would offset if not obviate the efficiencies sought in hiring the agent; and
- (2) Often workers are hired for specialized abilities that the principal does not possess and has only limited ability to understand and evaluate.

Because the principal cannot directly observe how much effort the agent is devoting to the task, he/she cannot compensate him/her on this basis directly. There is available to the principal, however, a “signal” of the agent’s effort level, which is complicated because it also includes a random component representing factors that are beyond the agent’s control (e.g. rainstorms that depress a salesperson’s performance in a given week). The random component of the signal has special distributional properties that ensure that unfavourable signal outcomes become less likely when the agent’s effort level increases.⁶ Hence, the outcome is statistically, although not directly, indicative of the agent’s effort level and can serve as the basis of a compensation contract.

There is also an assumption, sometimes implicit, that higher levels of effort by the agent will make more valuable work outputs more likely.⁷ An optimal compensation schedule based on the signal of agent effort can be derived from this set-up. The schedule maximizes expected profits for the principal and expected utility for the agent, subject to a constraint that ensures that the agent expects enough reward to secure him/her from the

⁵ The principal is assumed to be risk-neutral, or, sometimes, less risk-averse than the agent.

⁶ This “monotone likelihood ratio condition” is important for the model’s conclusions.

⁷ This assumption is sometimes not stated explicitly because the signal and work output are considered to be the same or at least strongly, positively correlated. The assumption is empirically wrong for many real settings. For example, Frederick Brooks in his classic, *The Mythical Man-Month*, described how increased effort in the form of more persons on the task degraded performance in creating the IBM 360 Operating System. Many athletic performances such as playing golf or hitting a baseball in clutch situations require relaxed reliance on autonomic processes: those who “try too hard” tend not to perform well.

labour market. Consideration of the properties of this schedule produces the following conclusions.

- (1) The agent demands additional compensation for bearing the probabilistic risk associated with the signal of his/her performance; thus, measuring and rewarding performance is more costly than if effort could be observed directly, because of the agent's risk aversion.
- (2) The outcome is *Pareto inferior* to the outcome that would result if direct observation of effort were possible; that is, the principal and agent would be collectively better off, and neither would be worse off, if effort were directly observable. Hence, there is economic value lost due to uncertainty in the signal of effort.
- (3) The resulting compensation will include a variable component that increases with the signal.

This last conclusion is frequently offered both as an explanation and an endorsement of increasing-in-output payment schedules (such as sales commission systems).

Within this framework, Banker and Datar (1989) broadened the discussion of observability to include the problems of the *precision* and *sensitivity* of the stochastic signal, which may complicate discernment of underlying agent effort levels, thereby frustrating the creation of effective compensation schemes. Precision, roughly defined, is the degree to which movement in the signal indicates movement in the underlying quantity of interest (e.g. effort) rather than the random disturbance. A very low-precision signal is not very useful because changes are too much due to random "noise".⁸ Sensitivity, roughly defined, is the degree to which a change in an underlying quantity of interest (e.g. effort) tends to change an available measurement indicator. A signal that is very low in sensitivity is not very useful because even large movements in the underlying quantity of interest result in small, difficult to discern movements in the signal. The fields of statistical process and quality control address related statistical issues involved in separating signal from noise in repetitive and stable industrial processes (see, for example, Ishikawa, 1985).

In contrast with the rigorous theorizing of agency economics, behavioural science has tended to approach measurement and observability via detailed empirical examinations. Early work in this area often focused on pathological phenomena. For example, Blau (1963) conducted field research on government bureaucracies and found that well-intentioned organizational

⁸ In such a situation, the agent will demand a very high risk premium to be exposed to the high level of uncertainty in the signal.

measurement programmes were consistently dysfunctional. In one study, agents at an employment office reacted to a system that measured job applicant interviews by shifting most of their efforts to interviewing, away from locating new job opportunities. Consequently, the office made fewer job placements.

Realizing that the system was dysfunctional, office managers enhanced the system to include eight measures, some of them ratios, such as “percentage of interviews that resulted in job referrals”. At first the enhanced system appeared effective, but a replication of the Blau study revealed that earlier dysfunctional behaviours had been replaced by more sophisticated dysfunctional behaviours. For example, agents “engaged in outright falsification . . . by destroying at the end of the day those interview slips that indicated that no referrals to jobs had taken place” (Blau, 1963, 50). Further attempts to fix this system by adding many more measures were to no avail; employees always adapted and dysfunction resulted.

In Blau’s view, the pattern of dysfunction that resulted from observability problems had three unfortunate characteristics. First, dysfunction seemed inherent in the attempt to measure organizational activity. Regardless of the incentive effect that the designers of a measurement system intended – regardless even of whether they meant to create any incentive effect at all – unintended incentive effects appeared. Second, dysfunction was persistent. It resisted efforts to eradicate it by adjusting the measurement technology. Finally, dysfunction was invariably hidden from the designers and users of the measurement system, often until some catastrophic failure occurred.⁹ The question “How do you know your system is not dysfunctional?” could not readily be answered. Ridgway (1956) reported a similar dysfunctional pattern in concluding that “quantitative performance measurements whether single, multiple, or composite . . . have undesirable consequences for overall organizational performance”.

This dysfunctional pattern has often been explained as resulting from *incompleteness* in measures (Lawler and Rhode, 1976). A measurement system suffers from incompleteness when measures do not capture all the *critical dimensions* of productive work. A dimension of activity is critical when no value can be produced without devoting some effort to the dimension. In Blau’s employment office example, searching for job opportunities was a critical

⁹ In many cases disguising dysfunction relied on explicit deception, but not always. Dysfunctional activities also took more subtle forms, such as overly optimistic reports or the convenient omission of detail.

dimension of the work being done by the office, but it was not being measured. Workers shifted their efforts from searching for job opportunities to conducting interviews, a dimension that *was* measured. The result was dysfunctional.¹⁰

The R–H model offers no explanation of the pattern of dysfunction that results from incompleteness in measures because it assumes implicitly that measures are complete.¹¹ In fact, the increasing-in-measured-output payment schedules that seem implicated in the pattern of dysfunction identified by Blau (and others) seem to be specifically recommended by the R–H model. The two streams of research are at odds on this point.

Economists eventually modified their models to address the problem of incompleteness and consequent dysfunction, bringing the two research streams closer together. Holmstrom and Milgrom (1991) modelled a situation in which the agent allocates effort across tasks, some of which are measurable and some of which are not. They show that, if one of the immeasurable tasks is critical to value creation and rewards increase with measurements on measurable tasks, then dysfunction will result. If the agent is willing to do valuable work without measurement-linked rewards, then letting him/her work for a flat fee is the best that can be done with incomplete measures.¹² Milgrom and Roberts (1992) concede that this result “imposes a serious constraint on the incentive compensation formulas that can be used in practice”. Holmstrom and Milgrom (1991) argue that the measurability of specific effort dimensions may be a determining factor in work efficiency, organization and organizational structure.

Knowledge work and observability

As Holmstrom and Milgrom (1991) have shown for general settings, if the problem of completeness cannot be solved then payment schedules that

¹⁰ The problem of incompleteness has been independently documented in numerous contexts leading up to the present, as the following examples show. Measuring police work by percentage of crimes solved (i.e. “clearance rates”) has been shown to result in some citizens’ complaints not being posted, in other complaints being posted only after they are solved and in minor crimes being worked on before major crimes (Skolnick, 1966). Measuring teachers on student test performance has been shown to result in narrowing of the educational mission to “teach the test” (Stake, 1971; Hannaway, 1992). Measuring tax collection rates has led US Internal Revenue Service employees to alter and falsify tax records.

¹¹ Specifically, the assumption that more valuable work outputs will be more likely when the agent increases his/her effort is indicated. The problem of incompleteness reveals that *how* effort is allocated matters at least as much as *how much* is allocated.

¹² Baker (1992) arrives at a similar set of conclusions via a somewhat different route.

increase with increases in observable dimensions of performance are dysfunctional. Because knowledge work tends to be oriented towards innovation and problem solving, it may benefit from efforts on dimensions that were unanticipated at the outset and for which viable measurement technology does not exist (either because the dimension is difficult to measure or because its importance was not recognized when the measurement system was designed). Knowledge work is multidimensional, and the criticality of dimensions may evolve dynamically. In work that is all about “smartness”, how smart someone works – that is, how cleverly one allocates effort to various tasks – is, necessarily, of importance. For these reasons (and others), the problem of completeness is especially prominent when the work being performed is knowledge work.¹³

Prescriptive works on the subject of organizational measurement often include recommendations that measures be as complete as possible. One popular treatment, for example, urges that organizational scorecards be “balanced” (Kaplan and Norton, 1992, 1996) by including non-financial as well as financial measures, in categories that more exhaustively cover what really matters to a business. Although balance might not be precisely the same thing as completeness, the evoked intuition is often the same. While recommendations of balance or completeness may be a step in the direction of improved practice, it remains important to understand when completeness can be achieved and in which work settings.

Observability issues that arise in physical settings have typically been framed in terms of asymmetry in *information* about workers’ hidden acts. Arguably, knowledge work generates more pronounced asymmetries, based not only on information asymmetries but also on knowledge (or even skill or talent) asymmetries between a manager and those he/she manages. A manager who has the same information as a worker can still lack the expertise needed to understand, attribute, evaluate and act on what he/she observes. For this reason, observability problems that afflict knowledge work may be particularly severe and persistent. A software company CEO we interviewed expressed this reality clearly: “it’s the real deep technical guys [who] are tough . . . This may be the most difficult issue we face . . . You have no idea what they’re doing. They sit there with 42 little windows open on their 17-inch monitor” (Austin and Larkey, 2000).

¹³ As we argue in the next section, knowledge workers do tend to be intrinsically motivated and their motivation can be relied upon and productively directed. This fact makes inability to solve the completeness problem less painful, and diminishes the relative importance of the issue of how *much* effort an agent is allocating.

It is important also to acknowledge that, like information asymmetries, knowledge asymmetries and resulting observability problems cannot necessarily be “designed away”. Many prescriptive treatments of organizational measurement focus almost exclusively on the measurement technology – on “choosing the right measures” – as if observability were entirely dependent on the shrewdness of measurement system designers. In fact, observability problems derive ultimately from the totality of the organizational situation and the specific setting of the measurement act, and not just from the measurement technology.

Persistent observability problems can be decomposed into three categories of more specific problems, all of which frequently appear in knowledge work settings, and any of which suffices to undermine performance measurement objectives.

- (1) *Measurability* is the degree to which important aspects of the observed work yield to cognitively simple and relatively compact quantification. If the activity in question cannot be measured in a way that is meaningful, with acceptable confidence in the validity of the measure, then measurement necessarily becomes more complicated. In such situations, less direct proxy measures are often sought, which creates potential for problems in the two categories that follow (below). Software quality, for example, is notoriously difficult to measure. Defect counts, a common measure, are at best only suggestive of quality as a customer defines it, in terms of suitability for the customer’s intended purpose. Similarly, human capital at a firm like Microsoft, where market value diverges dramatically from book value, would be difficult to quantify satisfactorily by using, say, counts of numbers of graduate degrees.
- (2) *Attributability* is the degree to which a measurement can be attributed to some causal object – an individual, a group or a process. Measuring something without knowing (or at least confidently hypothesizing) its relationship to a causal object is not very useful. Moreover, the ability to measure a thing does not ensure that the thing can be easily or usefully attributed to an actionable underlying cause.¹⁴ There tend, for example, to be important interdependencies in the production of knowledge work; it is rarely easy, even after the fact, to say who contributed what to the value in the final product.

¹⁴ Ishikawa (1985) has dealt in detail with the attribution problem in the context of quality control, in his distinction between “true quality characteristics” and “substitute quality characteristics”.

- (3) *Evaluability* is the extent to which the normative adequacy of a measurement can be judged. Measuring and attributing without knowing whether the measurement reflects favourably or unfavourably on the object of attribution is also not very useful. Standards and benchmarks typically assist with evaluability issues, but the ability to measure and attribute a thing does not guarantee that the thing can be easily or usefully evaluated. Much knowledge work results in products that are distinctive, and, in general, the greater the distinctiveness the lower the evaluability.

Prescriptive treatments and economic models have often assumed away these problems, especially the second and third. The R–H model, for example, deals with measurability in a limited way, in that the agent's effort expenditure cannot be observed directly. Because there is only one agent, however, there are no issues of attributability from interdependent production or complex production contexts. Evaluability is assumed, in that signal and work outputs are usefully related and effort allocation occurs along a single dimension. More effort is good; less effort is bad. Potential skill differences between agents, differences that might enable a more skillful agent to accomplish more with less effort, are effectively ignored in these models. Thus, R–H models misrepresent the production processes and managerial difficulties in knowledge work settings. Organizational and managerial advice derived from such misrepresentations may be unhelpful, even perverse. Much of the problem with R–H-like models applied to knowledge work settings is gross oversimplification. Knowledge work, at least the portion not automated from prior knowledge work, is intangible and conceptually complex; measurability and evaluability problems are common.

Useful proxy measures are not obvious when production may consist of staring vacantly out a window in deep thought or taking a Galton-like walk in an attempt to produce an unprecedented outcome such as the software for the next generation of military aircraft. Process proxies such as stare-minutes or kilometres walked are patently unsatisfactory. Proxies closer to outcomes are also difficult because (1) specifications for unprecedented systems evolve with science and experience; (2) there is no meaningful end point to production; the testing, evaluation, and modification cycle is perpetual for the life of the weapons system as users explore the effectively infinite space of operating contexts; and (3) by definition, there are no benchmarks for unprecedented systems. This is hardly day labourers moving rocks from a delivery point to a construction site, a production process more amenable to representation by R–H models.

In knowledge work, there are almost always vexing questions about the attributability of any chosen proxy measures. The collaborative and interdependent nature of much knowledge work makes it difficult, if not impossible, to understand measured outputs causally; the processes producing outputs tend to be evolutionary, idiosyncratic, and even erratic, bearing no resemblance to an interpretable experiment. Disputes about the parentage of knowledge accomplishments are commonplace with many claimants; knowledge failures are, as expected, orphans.

Measurability issues are a prerequisite for incompleteness. They necessitate the use of proxy measures in place of “true measures” – measures of what the organization truly values. Once proxy measures are introduced, questions about their relationship to true measures arise. Attributability concerns complicate attempts to establish persistent relationships between proxy and true measures, leaving workers, including managers, latitude to engage in dysfunctional behaviours. Problems of evaluability hinder the detection of dysfunctional behaviours as the pattern unfolds.

Several characteristics of organizational situations that have often been observed to complicate work measurement create particularly difficult problems for knowledge work measurement. Measures of knowledge work are more likely to be incomplete on account of the following factors.

- (1) *Context insensitivity*. Because measurement designers may lack vital knowledge about the work being done, contextual variables that have independent effects on proxy measures may be unknown or unknowable. Even if they are known, they may be ignored because of the conceptual complexity involved in defensibly measuring them. As has been argued, ways of affecting proxy measures that are known to workers but not to managers create potential for incompleteness.
- (2) *Inseparability*. Because of knowledge asymmetries and the highly collaborative nature of some knowledge work, measurement users may have difficulty in separating the work of one individual or group from that of another because the work is (or seems) so interdependent (Alchian and Demsetz, 1972). Consequently, workers may have the ability to move proxy measures in what seems like a favourable direction by shifting work difficulties onto other workers. Needless to say, a system that rewards such behaviours does not foster cooperation among workers.
- (3) *Reliability*. For knowledge work that is rapidly changing, it may become difficult to attribute, and evaluate because comparisons of a process from repetition to repetition are not obviously valid. The degree to

which measures succumb to problems of context insensitivity, inseparability, etc. may change from repetition to repetition. Empirically discovered correlations between measured quantities and underlying phenomena may not persist. If the measurement process is a statistical one, the problem manifests itself as a fundamental incoherence in the population definition.

- (4) *Ex post causal ambiguity*. Knowledge asymmetries provide latitude for variation in *ex post* interpretations of events. If a complex computer error shuts down a production facility, for example, there may be only two or three people who truly understand what has happened. The parties involved may take advantage of this fact by lobbying for specific interpretations of measurements and events. This practice has a destabilizing effect on efforts to establish underlying causal attributions. In established or more physical measurement settings, there is often consensus on the causal models that underlie measurement. In many knowledge work settings, no such consensus exists. Moreover, the extent to which such a consensus is realizable may be limited by the rapidly changing nature of much of knowledge work.

The problem of motivation

Holmstrom and Milgrom's (1991) assumption that the agent might do some valuable work when paid only a flat fee because he/she is somehow internally¹⁵ motivated to exert effort on another's behalf, is not remarkable in the context of the behavioural literature, but it is non-standard in economic theory. Holmstrom and Milgrom do not explore other possible questions related to this kind of motivation, such as "How might internal motivation be used to produce more value?" or "In what conditions might internal motivation be a viable control mechanism?". Behavioural scientists, however, have studied human motivation extensively and the conditions in which external and internal motivations can be used for organizational control (e.g. Amabile, 1996; Eisenhardt, 1985; Deci and Ryan, 1985; Ouchi, 1979, 1981; Vroom and Deci, 1970; Thompson, 1967; McGregor, 1960; March and Simon, 1958).

¹⁵ We use the expressions "internal" and "external" (or "intrinsic" and "extrinsic") to describe a specific distinction between types of motivation. External motivations are those that are linked to objectively discernible measured outcomes. Internal motivations are those for which the normative adequacy is determined subjectively by the individual whose performance is under consideration.

Thompson (1967) distinguishes between behaviour- and outcome-based measurement for the purposes of control. To the extent that behaviour is observable, he argues, measures will be based on behaviour. Similarly, outcome measures will be used when desired outcomes can be observed readily. Ouchi (1979) argues that, when neither behaviour nor outcomes are very observable, then socialization or “clan mechanisms” will be the only recourse. These mechanisms achieve control by increasing the congruence of objectives between the organization and individuals. The parallel to the Holmstrom and Milgrom (1991) model, in which value is produced without external reward via objectives that are assumed to be at least somewhat congruent, is apparent. More recent work by Holmstrom and Milgrom (1994) has taken an approach similar to that of Ouchi (1979) in suggesting that the choice between the organizational and market coordination of production is contingent on measurability/observability conditions. A remaining difference between the two research streams, however, is in the degree to which internal motivations are assumed to be a viable means to control. For the most part, economics continues to treat the preferences of principal and agent as opposed and exogenous, while behavioural research has a long-standing tradition of exploring ways of usefully altering worker preferences (e.g. Barnard, 1938; McGregor, 1960; Ouchi, 1979).¹⁶

Some researchers have argued that not only are worker preferences alterable but also that external motivation schemes themselves alter preferences with a detrimental effect (e.g. Amabile, 1996; Frey, 1993; Kohn, 1993; Deming, 1986; Deci and Ryan, 1985; McGraw, 1978). External motivations have been shown to “crowd out” internal motivations. Once offered rewards for taking specific actions, workers become unwilling to take actions that are not connected with specific rewards; the external reward system shifts the locus of control from worker to the external system, thus reducing the agent’s sense of self-determination. Moreover, an offer of external reward for actions that would have been taken because of internal motivation can have an

¹⁶ Leavitt and March (1988) summarize this difference between the two literatures in explaining Barnard’s views on altering worker preferences: “In modern terms, Barnard proposed that an executive create and sustain a culture of beliefs and values that would support cooperation. The appeal is not to exchanges, Pareto optimality, or search for incentive schemes; it is to the construction of a moral order in which individual participants act in that name of the institution – not because it is in their self-interest to do so, but because they identify with the institution and are prepared to sacrifice some aspects of themselves for it.”

insulting or demeaning effect (Hirsch, 1976); the offer of a reward when none is required creates cognitive dissonance (Festinger, 1957).

Amabile (1996) cites a large body of research on the social conditions favourable to creative work in arriving at what she proposes as a general principle: “Intrinsic motivation is conducive to creativity, but extrinsic motivation is detrimental” (p. 15). She argues also that, “intrinsic motivation appears to be essential for high levels of creativity” (p. 17). Laboratory experiments, whether they involve people considered creative experts or novices, show consistently that work outcomes are judged less creative when efforts are linked with extrinsic rewards. A theory proposed by McGraw (1978) predicts that extrinsic motivation will enhance performance on “algorithmic” tasks, but undermine performance on “heuristic” tasks. Algorithmic tasks are those for which the path to a solution is clear and straightforward – tasks for which an algorithm exists. Heuristic tasks are those not having a readily identifiable path to a solution – tasks for which an algorithm must be invented (Hilgard and Bower, 1975). Knowledge work may, of course, contain both heuristic and algorithmic elements. The degree to which a given knowledge work job is composed primarily of algorithmic or heuristic tasks may then greatly determine whether the motivational effects of performance measurement will be helpful or harmful.

In apparent contradiction to the research on negative effects of measurement on motivation, Simons (1996) points out that there do appear to be successful organizational control systems that combine external rewards and socialization in a way that facilitates particularly effective responses to environments in which observability problems are strongly present. These systems, which he terms “*interactive control systems*” (ICSs), not only motivate employees but also serve as the primary mechanism for directing discussion around issues that are vital to the organization’s survival.

ICSs embed measurement activities in intensive interpersonal processes. In a Harvard Business School case that illustrates an ICS in action, Codman and Shurtlef, a division of Johnson and Johnson, uses measurement intensively but invariably within the context of lengthy meetings between managers and those being managed. Underlying measurement activities are a set of norms about appropriate behaviours with respect to measurement uses that are widely known and accepted. Embedding measurement into social interaction in this way apparently engages the clan mechanisms referred to by Ouchi (1979).

Why the use of ICSs does not result in crowding out behaviours is not entirely clear. The socialization that occurs with successful ICSs may

somehow prevent shifting the locus of control from workers to the external system. In the Codman and Shurtlef case, the norm that places the right to change measurement targets solely with the person responsible for meeting the target is, for example, suggestive in this regard. The involvement of workers in the definition and operation of a measurement system may mitigate the potential “disempowering” effects of defining performance externally. Similarly, socialization may also sometimes, by a mechanism as yet unknown, diffuse the cognitive dissonance that comes from offering rewards for accomplishments that were internally motivated. Alternatively, and more pessimistically, ICSs that seem successful may simply be systems that have not yet revealed their dysfunctional tendencies.

Also unclear is the more general relationship between external factors that influence or constrain work and intrinsic motivation in creative tasks. Although the social psychological results that demonstrate negative effects of extrinsic reward on the creativity seem robust, it is obvious that creative workers work successfully within external constraints that are at least loosely linked with rewards. For example, a successful theatre company always works within a firm deadline (opening night) and the penalties for failing to meet the deadline are severe. But for some reason, such external constraints seem to have different effects than performance measurement with designed links to external penalty or reward. Indeed, some creative workers describe harsh constraints as an important stimulus for breakthrough creativity (see, for example, Austin and Devin, 2003). Amabile (2005) has conjectured that the difference may partly be accounted for by the extent to which workers “internalize” an external constraint. If creative workers accept an external constraint and make it their own, perhaps even wear it as a badge of honour, it may have no detrimental effect on creativity, and may instead strengthen intrinsic motivation. ICSs described by Simons appear to include mechanisms for encouraging this kind of internalizing of constraints.

Knowledge work and motivation

The prominence of the problem of incompleteness in measures of knowledge work has implications for motivation to which we have already alluded. In the presence of incomplete measures, Holmstrom and Milgrom (1991) suggest paying the agent a flat fee and relying on any partial congruence between the objectives of principal and agent to produce value. If we assume

that incompleteness of measures is the norm for knowledge work, then we are forced into consideration of internal motivation as a primary means of directing and controlling agent action. Methods of engaging clan mechanisms (Ouchi, 1979) or ICSs (Simons, 1996) become extremely important. The likelihood of creating successful systems that rely on external reward and compliance with plans is reduced, not only because performance verification is frustrated by incompleteness in measures but also because plans may be unstable in knowledge work settings.

This news is not so bad as it might seem, however, because it is commonly observed that knowledge workers are highly motivated. Knowledge workers tend to self-select into their professions because they like the work. Obtaining effort from such workers is less an issue than directing their effort. The problem of objective congruence is not so much about workers' aversion to effort or risk as it is about independently minded workers whose performance cannot be verified having different ideas about what ought to be done.

The previously mentioned CEO of a software company recently related the problem of managing one of her most talented developers. This developer, in addition to taking a strong interest in the success of the company, was also a "free software" activist, ideologically opposed to software patents and the ownership of ideas as expressed in software. At a time when the company was working under deadline for a client, struggling for its very life, this developer was working on a programme to generate random patent applications, ostensibly to frustrate the patent authorities. The developer also finished work for the client on time, but his "extracurricular" activities were non-optimal (to say the least) from the perspective of the management team. This same CEO, however, conceded that there was little to be done about this problem, and that such knowledge workers were uncontrollable in any traditional sense. Ivan Sutherland, manager of some of the researchers who created the Internet, says of such workers, "you can maybe convince them that something's of interest and importance, but you cannot tell them what to do" (Hiltzik, 1999).

In knowledge work, selecting appropriately skilled workers may be much more important than directing and controlling them once selected; the programmer who can perform a critical task with five per cent of his effort over some period is always preferable to the programmer who cannot perform the same task with 100 per cent of his effort over the same period.

The potential for crowding out is a particular concern for knowledge work. If external motivation crowds out internal motivation, actually

reducing workers' willingness to do more than exactly what they are paid to do, then it reduces the tendencies towards self-direction and initiative that are crucial in knowledge work. Furthermore, external reward systems may interfere with communication by encouraging workers to censor and adjust information flows to managers. Workers worried about how their measures "look" may not share information as willingly or effectively. This is likely to present a significant problem, since it has often been argued (e.g. Eisenhardt and Brown, 1997; Leonard-Barton, 1995) that effective communication is a key element of success in knowledge work, which requires the coordination of concepts that are to a great extent only tacit in the work environment.

The problem of TASK differentials

Like performance measurement, the subject of talent, skill and knowledge differentials draws historically from numerous fields. Differential capabilities in problem solving have, for example, been treated extensively in developmental and cognitive psychology (see, for example, Simon, 1989). The field sometimes referred to as "complex adaptive systems" has explored the manner in which algorithmic behaviours of organisms or automata evolve differentially in response to environmental conditions. Economists have dealt with TASK differentials as the basis for adverse selection problems; differentials are assumed to be exogenous and the challenge becomes distinguishing the level of capability in agents being selected.

However, as we have observed elsewhere (Larkey *et al.*, 1997), there has been surprisingly little substantial research into the nature of differential inherent capabilities. The differential capabilities in playing basketball between Michael Jordan and the average college faculty member would be the primary determinants of the outcome of a contest between the two. Yet we lack, for the most part, any theoretical framework that would predict or account *ex post* for the outcome. To some extent, we lack this theory because capability differentials are context-specific. Skill in basketball is specific to basketball. Unless the context is worth researching on its own (as, for example, organizational capability to innovate is – see, for example, Leonard-Barton, 1995), then research into skill lacks the generalizability that researchers value in their products.

The absence of theoretical representations of skill becomes very relevant to performance measurement, in that much of its underlying analysis,

particularly in economics, is game-theoretic in nature. Agency theory is based on the notion that equilibria will emerge as the result of mutual optimization by principal and agent. There is an implicit assumption that all participants in the game can and will optimize. Capability differentials are not present in this representation at all. This is unfortunate for contexts in which these differentials are primary determinants of outcomes (think of our basketball context between Michael Jordan and an average faculty member). Binmore (1990) provides a sweeping, critical review of game theory and concludes that the absence of specific representations of players' abilities and tendencies is a glaring hole in the applicability of game and economic theories. Similarly, it poses a challenge to the relevance of performance measurement.

Knowledge work and TASK differentials

Curtis, Krasner and Iscoe (1988) have documented tremendous disparities in the productivity of software developers. It has become a common belief in the knowledge work fields related to high technology that the inherent capability of individual workers is one of the most important factors in the success of an organizational initiative. The chief information officer (CIO) of a major Silicon Valley hardware manufacturer estimated recently that there were only 1,000 or so world-class software developers on the planet, and stated that his firm was dedicated to holding on to theirs and acquiring more. Increasingly, companies such as Trilogy Software, in Austin, Texas, fashion their competitive strategies around acquiring the very best technical talent emerging from universities and then facilitating their work with a minimum of obtrusive structure (see Austin, 1998). Top-tier consulting firms have long used this strategy, as do universities in hiring and managing the development of faculty.¹⁷

Conventional performance measurement frameworks have very little to say about managing TASK differentials and can often have dysfunctional effects in contexts in which TASK differentials are known to be important. Understanding why this is so requires examining the intended uses of

¹⁷ V. G. Narayanan of the Harvard Business School has observed (in a private communication) that the job markets in such settings do not "clear" in the economic sense. In markets in which top talent is regarded as the determining factor, firms battle furiously for the top-quality workers and refuse to extend offers to those who are not perceived as being of this quality.

performance measurement in organizational settings. The intended uses of measurement can usefully be divided into two categories.

- (1) *Motivational measurement* is explicitly intended to affect the people who are being measured. An example of measurement in this category is sales tracking linked to a sales commission system. Used in this way, measurement is an attempt to control individual activity that, it is assumed, will not be congruent with organizational objectives in the absence of the measurement. This is the use of measurement that is implicit in most conventional measurement frameworks.
- (2) *Informational measurement* is valued primarily for the logistical, status and research information it conveys, which provides insights, supports organizational learning and allows better short-term management and long-term improvement of organizational processes. An example of measurement in this category is data gathering for the purpose of understanding how to redesign a business process. This use of measurement has little to do with control and much to do with learning.

In knowledge work settings, motivational uses of measurement are unlikely to be helpful, for many reasons. Incompleteness in measures, if present, will probably result in a distortion of effort allocations that is not constructive. The silver lining in all this, however, is that the informational uses of measurement are likely to be very helpful in knowledge work settings. Knowledge workers have an appetite for measurement information that can help them do their jobs or improve their own performance. The idiosyncratic resourcefulness that knowledge work requires can be greatly enhanced by effective use of informational measurement, much as a sprinter's performance can be enhanced through training with a stopwatch.

It would seem to follow that organizations engaged in knowledge work would be well served by instituting purely informational measurement systems and avoiding motivational uses of measurement. Avoiding motivational responses so as to preserve the validity of informational measurement can be very difficult, however. Because the distinction between the two categories of measurement use is not inherent in the information itself but, rather, in how the information is used, it is nearly impossible to reassure workers credibly that the purpose of a measurement system is purely informational. The transformation of an informational system of measurement to a motivational one can be triggered by seemingly minor events.

In an interview from an earlier study (Austin, 1996), one subject described an organization engaged in knowledge work in which workers were

measuring their own job processes, posting and comparing measurement information and using it to refine processes and improve performance:

Then one day, a very high-level president . . . was taken through the hall . . . [and] he saw [measurement information posted on a wall] and said “What’s this all about?” They told him and he said “This is wonderful.” He took out a red pen . . . and circled [one group’s numbers] . . . and he wrote “great work” and signed his name . . . He put the pen away and walked off. Literally the next day, the graphs came off the walls. No one ever put any graphs up again.

It is worth noticing that the act that caused the system to be abandoned was not a punishment but a reward. The red pen reward, although well intentioned, changed a system conceived to facilitate self-directed work (an informational system) into an explicit external reward mechanism (a motivational system), and, in the process, destroyed the system.

Given that informational measurement systems can so easily become motivational, it would seem difficult to avoid the dysfunctional effects of performance measurement in knowledge work settings. That there is hope, however, is demonstrated by the fact that the measurement system in place in the above “red pen” example did work well until the intervention by senior management. Through the use of clan mechanisms (Ouchi, 1979), ICSs (Simons, 1996) or some other means, it may be possible for measurers with informational aims to diffuse unintended motivational effects.

There is some irony here, in that the incentive design problem is turned on its head relative to its formulation in agency economics. Rather than *defining* explicit incentives that are expected to be the only way of productively motivating self-interested agents, the informational measurement challenge is to *diffuse* implicit incentives that might tempt highly capable self-interested agents to exploit imperfections in the measurement system. The principal tools to accomplish this are instruments of socialization, aimed at modifying worker preferences. Explicit incentives, such as the one introduced by the red pen, encourage reductions or falsifications in vital communication.

Research challenges

The challenges for research in the area of knowledge work extend naturally from its distinctive characteristics.

- Research frameworks must emphasize the completeness problem and explore the potential for the problems of measurability, attributability

and evaluability to complicate and make dysfunctional any attempts at performance measurement. Knowledge work is primarily about “working smart”. The research on measuring knowledge work should not, therefore, be primarily (if at all) about working hard.

- Because of the prevalence of the problem of incompleteness of measures in knowledge work, the degree to which workers can be internally motivated to create value must be endogenously included in performance measurement models. For knowledge work, it is most likely that marginal returns from investments in improving worker conditions, motivations or skills are greater than marginal returns from investments in performance verification. The latter sort of investment is the focus of much of the traditional research in this area. In practice, many of the best companies have learned that no dress code, pet-friendly work environments, an accessible swimming pool or volleyball court, and free caffeinated beverages do much more to increase output than bureaucratic scoring systems, however clever those systems may be.
- TASK differentials must be included somehow in the representation of measurement and behaviour. Success in knowledge work has more to do with what workers are capable of than with how hard workers are working. Differential capabilities are not present at all in most performance measurement frameworks, but must be introduced.
- The importance of self-direction in knowledge work, and the tendency of knowledge workers to produce value by engaging in non-repetitive, non-standard, and innovative activities must be acknowledged in structuring work processes and management systems. This may be the most profound challenge at the heart of managing knowledge work. When work processes vary in ways that cannot be anticipated or prescribed in advance, and when it is exactly those variations that contain the promise for the most value creation, how should measurement function? Surely not to enforce conformance to a prescribed process. What a classic industrial measurement system might classify as a “quality problem” – an exceptional event in the execution of a standard process – might qualify instead in a knowledge work setting, as an “important innovation.” To the extent that performance measurement constrains appropriate innovation, it may not serve the purposes of the organization. The already cited description by a software company CEO of the difficulties of managing knowledge workers concluded by making this point: “When [your business shifts]

you'll often find the seed for the shift in that group [of highly capable knowledge workers] because they're not really paying attention to you all along anyway. They were worried about some way-out-there trend. They'll see it and there will be something there. [The key to] how to manage change is in that group of folks you don't have a lot of control over" (Austin and Larkey, 2000).

Practical implications

As concerns performance measurement, the primary practical implications of the distinctive characteristics of knowledge work have to do with the very different ways that measurement interacts with work in the knowledge work setting. If the collaborative organizational model was "organization as machine with interchangeable parts", the knowledge work model is "organization as collaborative ensemble". The parts are selected for their roles and are not interchangeable. Talent, skills and knowledge are too individual and knowledge is too often tacit to accommodate traditional notions of performance comparability.

The presence of TASK disparities necessitates a shift from traditional feedback-based control to *measurement-facilitated discovery*. Of vital importance in using measurement in the latter sense is that workers, not managers, become the primary consumers of the information. Managers give up their roles as primary keepers and analysers of measurements. While this has been advocated as a desirable arrangement in many physical work settings, knowledge work settings may offer no alternative (other than dysfunction).

The most important duty that managers have under this model of measurement is to convince workers to participate willingly in the measurement and distribution of information that could potentially be used against them. We consider this a fundamental principle of measuring knowledge work: *that it requires that workers engage in a behaviour that is fundamentally irrational on a self-interested individual basis*. Every day knowledge workers are confronted with choices between actions that will make them "look good", according to incomplete performance measurements, and other actions that will have a less favourable impact on measured performance but will, in fact, create more value. The success for performance measurement in knowledge work settings hinges on which choice the worker consistently makes.

References

- Amabile, Teresa M. (2005). Conversation with the author.
- Amabile, Teresa M. (1996). *Creativity in Context*, Westview Press, Boulder, Colorado.
- Austin, R. D. (1999). *Trilogy Software, Inc.*, Case no. N9-699-034. Boston: Harvard Business School Press.
- (1996). *Measuring and Managing Performance in Organizations*. New York: Dorset House.
- Austin, Robert D. and Patrick D. Larkey (2000). "Performance Based Incentives in Knowledge Work: Are Agency Models Relevant?" *Int. J. Business Performance Management*, Vol. 2, Nos. 1/2/3, p. 57.
- Alchian, A., and Demsetz, H. (1972). Production, information costs and economic organization. *American Economic Review*, 62(5), 777-95.
- Baker, G. P. (1992). Incentives, contracts and performance measurement. *Journal of Political Economy*, 100(3), 598-614.
- Banker, R. D., and Datar, S. M. (1989). Sensitivity, precision and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, 27(1), 21-39.
- Barnard, C. I. (1938). *The Functions of the Executive*. Cambridge, MA: Harvard University Press.
- Binmore, K. (1990). *Essays on the Foundations of Game Theory*. Cambridge: Basil Blackwell.
- Blau, P. M. (1963). *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*. Chicago: University of Chicago Press.
- Brooks, Frederick (1995). *The Mythical Man-Month: Essays on Software Engineering*, 20th Anniversary Edition Addison-Wesley, Reading, MA.
- Curtis, B., Krasner, H., and Iscoe, N. (1988). A field study of the software design process for large systems. *Communications of the ACM*, 31(11), 1268-87.
- Davenport, Thomas, H. (2005). *Thinking for a living: How to Get Better Performances and Results from Knowledge Workers*, Harvard Business School Press, Boston, MA.
- Davenport, T. H., and Prusak, L. (1998). *Working Knowledge: How Organizations Manage what They Know*. Boston: Harvard Business School Press.
- Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and Self-determination in Human Behavior*. New York: Plenum Press.
- Deming, W. E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Press.
- Eisenhardt, K. M. (1985). Control: organizational and economic approaches. *Management Science*, 31(2), 134-49.
- Eisenhardt, K. M., and Brown, S. L. (1997). *Competing on the Edge*. Boston: Harvard Business School Press.
- Festinger, L. A. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.
- Frey, B. S. (1993). Motivation as a limit to pricing. *Journal of Economic Psychology*, 14, 635-64.
- Hannaway, J. (1992). Higher-order thinking, job design and incentives: an analysis and proposal. *American Education Research Journal*, 29(1), 3-21.
- Hilgard, E. and Bower, G. (1975). *Theories of Learning* (4th ed), Prentice-Hall, Englewood Cliffs, NJ.

- Hiltzik, Michael A. (1999). *Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age*, New York: Harper Collins. pp. 145–146.
- Hirsch, F. (1976). *The Social Limits to Growth*. Cambridge, MA: Harvard University Press.
- Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10, 74–91.
- Holmstrom, B., and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55, 302–28.
- (1991). Multi-task principal–agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organizations*, 7, 25–52.
- Ishikawa, K. (1985). *What is Total Quality Control? The Japanese Way* (trans. Lu, D.J.). Englewood Cliffs, NJ: Prentice-Hall.
- Jensen, M. C., and Meckling, W. H. (1976). Theory of the firm: managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics*, 3, 305–60.
- Kaplan, R. S., and Norton, D. P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.
- (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Kohn, A. (1993). Why work incentives fall down on the job. *USA Today*, 16 December, A13.
- Larkey, P., Kadane, J., Austin, R., and Zamir, S. (1997). Skill in games. *Management Science*, 43(5), 596–609.
- Lawler, E. E. III, and Rhode, J. G. (1976). *Information and Control in Organizations*. Santa Monica, CA: Goodyear.
- Leavitt, B., and March, J. G. (1988). Chester I. Barnard and the intelligence of learning. *Annual Review of Sociology*, 14, 319–40.
- Leonard-Barton, D. A. (1995). *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*. Boston: Harvard Business School Press.
- March, J. G., and Simon, H. A. (1958). *Organizations*. New York: John Wiley.
- McGraw, K. O. (1978). The detrimental effects of reward on performance: a literature review and a prediction model. In Lepper, M. R., and Greene, D. (eds.), *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*, 33–60. Hillsdale, NJ: Erlbaum.
- McGregor, D. (1960). *The Human Side of the Enterprise*. New York: McGraw-Hill.
- Milgrom, P., and Roberts, J. (1992). *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Nonaka, I., and Takeuchi, H. (1995). *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford: Oxford University Press.
- Ouchi, W. G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9), 833–48.
- (1981). *How American Business can Meet the Japanese Challenge*. Reading, MA: Addison-Wesley.
- Prahalad, C., and Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, 68(3), 79–91.
- Ridgway, V. F. (1956). Dysfunctional consequences of performance measurement. *Administrative Science Quarterly*, 1(2), 240–7.
- Ross, S. A. (1973). The economic theory of agency: the principal's problem. *American Economic Review*, 63(2), 134–9.

- Simon, H. A. (1989). *Models of Thought*. New Haven, CT: Yale University Press.
- Simons, R. (1996). *How Managers Use Innovative Control Systems to Drive Strategic Renewal*. Boston: Harvard Business School Press.
- Skolnick, J. (1966). *Justice without Trial: Law Enforcement in Democratic Society*. New York: John Wiley.
- Stake, R.E. (1971). Testing hazards in performance contracting. *Phi Delta Kappan*, 54, 583–9.
- Thompson, J. D. (1967). *Organizations in Action*. New York: McGraw-Hill.
- Vroom, V. H., and Deci, E. L. (eds.) (1970). *Management and Motivation*. Harmondsworth: Penguin.

Riitta Katila

Introduction

Patents and patent citations are increasingly used as measures of innovation performance (e.g. Katila, 2002). However, confusion exists over the applicability of these measures and over the appropriate patent citation lag to be used. This chapter examines the measurement of innovation performance through patents, focusing especially on how to measure the radicality¹ of innovations by using patent data.

The existing literature provides a wide array of definitions of “radical” innovations. In this contribution I propose that the previous definitions of radicality can be arranged into four broad categories – industry-, organization-, user- and technologically radical – each addressing a different dimension of radicality (see table 14.1). The first category of radical innovations defines “radical” as new or disruptive to the *industry*. Radical new products at the level of the industry dominate and make obsolete the previous products in established markets, can give rise to new industrial sectors (Achilladelis, Schwarzkopf and Cenes, 1990) and affect the market power relations in the industry (Henderson, 1993).

The second category of radical innovations defines radical as new to the *organization*. Organizationally radical innovation may be defined as innovation that incorporates, for example, a technology that is new to the firm but may be well understood by others (Green, Gavin and Aiman-Smith, 1995). Organizational radicality has also been described as a degree of change the innovation makes in the existing practices of the organization. The third category defines radical as new to the *users*. User-radical innovations fulfil customer needs much better than the existing products, but may require new

¹ Radical innovations have also been called path-breaking, discontinuous, revolutionary, new, original, pioneering, basic or major innovations (Green, Gavin and Aiman-Smith, 1995).

Table 14.1: Definitions of radicality of innovation

<i>Characteristics of each level of analysis</i>	
Industry	<ul style="list-style-type: none"> • Disruptive to the existing players of the industry • Requires incumbents to acquire new skills, practices and patterns of thinking
Organization	<ul style="list-style-type: none"> • Disruptive to the organization • Requires organization to acquire new skills – incompatible with existing practices
User	<ul style="list-style-type: none"> • Possibly disruptive and new to the users • Provides advantages over prior product forms, but also requires new skills from the user
Technology	<ul style="list-style-type: none"> • Disruptive to the experts of current technology • Requires new scientific and engineering knowledge and new problem-solving approaches

skills to be used. Utterback (1994) defines radical innovation as one that has potential for delivering dramatically better product performance.

The fourth category of radical innovations defines radical as *technologically* new and significant. Technologically radical innovations include new know-how, a new set of engineering and scientific principles or a new problem-solving approach (Henderson and Clark, 1990). They require actors within a technological area to develop qualitatively new technological capabilities and incorporate technology that is a significant departure from existing technology (Henderson and Clark, 1990). This chapter focuses on technologically radical innovations and uses patent data to measure them.

I begin with the motivations for measuring innovation performance and its radicality, and discuss prior work in the area. I then present a study of biotechnology companies in which patent-based measures of radicality are applied. The results of this study are further analysed from the point of view of radical innovation measurement. Based on this analysis, several recommendations for future work are given. For example, researchers using patent-based measures are urged to check the validity of their results by experimenting with several different citation lags. Implications for theory and practice conclude the contribution.

Radical innovation performance

Why is radical innovation measurement important?

Studying the measurement of radical innovations is important for at least three reasons. First, studies in the technology management literature propose that radical innovations increase firm performance and competitive

advantage. For example, industry leaders often produce incremental improvements and follow their core technologies to obsolescence and obscurity, while companies that are able to produce technologically radical innovations become the new leaders (Mitchell, 1989). Consequently, the accurate and objective measurement of radical innovation is important for managers as well as for performance researchers.

Second, despite the importance of radical innovation, and its wide use in the literature, there is relatively little work on how to measure radical innovation. For example, the different levels of analysis by which radicality can be defined, and the continuous nature of radicality, are poorly understood (Green, Gavin and Aiman-Smith, 1995). There is a need for more work to help firms measure the different types of radicality that result from their innovative efforts.

Third, the study of the measurement of radicality is motivated by the observation that the concept of radical innovation is increasingly being used in new research areas. Previously, radical innovation has been a widely used construct in industrial organization economics and strategy studies. The general conclusion of this work is that incumbents have a somewhat reduced incentive to innovate radically because of their existing interests in the technology and market (Henderson, 1993). Recently, researchers on collaborative relations and networks, for example, have started to ask how the incumbents' lack of incentives to innovate radically affects the outcomes of collaborative relationships that the incumbents are involved in. In all, both the theoretical and the practical importance of radical innovation motivates the study of reliable and valid operationalizations of this construct.

Measurement of radical innovation

Although radical innovations have been operationalized using several methods, there is no commonly accepted way to measure the radicality of innovation. In this subsection several measures used in prior work are discussed. I discuss the strengths and weaknesses of these measures, and specifically point out how operationalizations using patent data can address some of the central weaknesses of the other measures.

Radical innovation has been measured using a variety of methods. Some authors have used qualitative methods such as expert or manager interviews to determine the most radical innovations in the industry (see, for example, Achilladelis, Schwarzkopf and Cenes, 1990, Green, Gavin and Aiman-Smith, 1995, and Henderson, 1993). Other researchers combine

qualitative measurement with quantitative data. Anderson and Tushman (1990) operationalize a radical design as an innovation that improves the product's current performance "frontier" by a significant amount – for example, a significant improvement in the central processing unit speed of a computer. Christensen and Rosenbloom (1995) use performance improvement data to operationalize radical innovation.

The above-mentioned operationalizations of radical innovation have three main weaknesses. First, many studies use a binary categorization of radicality: innovations are either radical or incremental. This categorization, however, does not necessarily correspond to the more fine-tuned reality: radicality is a continuum. Second, in many studies the evaluation of radicality is based on subjective and potentially partial assessment by managers, industry experts or customers. The reliability of these measures can be context-dependent (Pavitt, 1988). Third, prior operationalizations rarely distinguish between the four types of radicality discussed earlier in this chapter. Drawing definitive theoretical conclusions based on studies that fail to make these distinctions is difficult.

The three weaknesses of radical innovation measures discussed above can, however, be addressed by using patent-based measures of radicality. Patents and their subsequent citations provide a *continuous* and a relatively *objective* measure of technologically radical innovations. By definition, patents include *technologically novel*, useful and distinct knowledge (Walker, 1995) and thus provide a good measure of technological radicality as defined above. Several studies in recent years have used patents as a measure of innovation performance (e.g. Dutta and Weiss, 1997; Henderson and Cockburn, 1994; Jaffe, Trajtenberg and Henderson, 1993).

In addition to these methodological strengths of patent-based measures, two other factors also motivate the use of patent-based measures of innovation performance. First, the use of patenting is increasing. Kortum and Lerner (1999) document a surge of approved patents in recent years in many high-technology industries, and Arora and Gambardella (1994) further argue that the importance of patents as innovation appropriability mechanisms looks set to continue increasing in many industries in the future, as several technological disciplines become more universal and the knowledge becomes easier to articulate for patenting. Second, easier electronic access to patent data through, for example, European and US patent office databases has increased the possibilities for using patent data. Thus, patents are even more likely to be used as measures of innovation performance in the future in R&D organizations as well as in innovation research.

Despite the above-mentioned benefits of patent-based operationalizations, these measures also, inevitably, have limitations (see also Walker, 1995). Previous studies have pointed out that the propensity to patent varies considerably across industries, and therefore patents provide the best, comparable measure of innovation when the analysis is restricted to one industry (Ahuja, 1996). It is also the case that patents can measure only a fraction of the total research output, but, nevertheless, they seem to “provide one of the few direct quantitative glimpses into the innovation process available to us” (Griliches, 1984, 14).

Another issue concerning the use of patent measures is the differences among firms in their propensity to patent (Pavitt, 1988). One potential concern is the tendency of some companies to abuse the patent system, such as patenting only for the purpose of suppressing or preventing competition. However, patenting and renewal fees are designed to prevent this type of abuse (Walker, 1995). Finally, patents can vary in their qualitative importance. In recent years several researchers have proposed ways to distinguish and measure these qualitative differences. One of these methods, citation weighting, is discussed below.

Citation-weighted patents as a measure of the radicality of innovation

Several authors (such as, for example, Jaffe, Trajtenberg and Henderson, 1993, and Trajtenberg, 1990) have argued that patents can vary enormously in their importance and value, and simple patent counts are thus unlikely to capture the qualitative differences in innovative output totally. Consequently, Henderson and Cockburn (1994) use granting in two of the three major geographical markets as an indicator of the patent’s importance. Other authors have added citations to patent counts – i.e. citations that the focal patent receives in subsequent patents – to measure the radicality of a particular innovation (e.g. Dutta and Weiss, 1997). Receiving several citations – i.e. being the basis for a number of subsequent innovations – is proposed as an indication of the technological significance of an innovation. The argument is that citation-weighted measures indicate the technological as well as the economic value of the innovation better than simple patent counts (Albert *et al.*, 1991; Dutta and Weiss, 1997).

Trajtenberg (1990) demonstrates that citation-weighted patents are a valid measure of radical innovation by confirming a significant relationship between citation-weighted patent counts and independent measures of the economic and social value of the same patents. Furthermore, Carpenter,

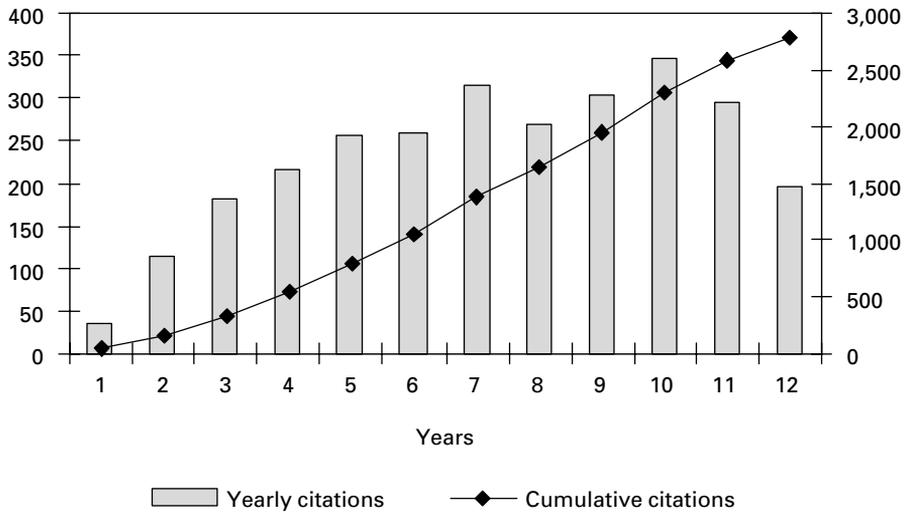


Figure 14.1: Average citation pattern for 242 US biotechnology patents

Notes: Patents applied in 1980–7. Patents citing the focal patents up to twelve years after the application of each of the patents are illustrated. Self-citations are excluded.

Narin and Wolf (1981) show that patents that resulted in radical innovations (100 awarded products) received more than twice as many citations as a matching sample of random patents. Finally, Dutta and Weiss (1997) used citation-weighted patent data to discriminate between technologically significant and incremental innovations.

Figure 14.1 provides an example of citation-weighted patent analysis. This figure shows an average citation pattern for 242 biotechnology patents. As illustrated by this figure, citation data provide two main sources of information: the number of citations that each patent receives and the timing of these citations. In these data the 242 patents receive cumulatively almost 3,000 citations in subsequent patents. In addition, two citation peaks are identified: first, the biotechnology patents are cited increasingly until seven years after their application; then, somewhat surprisingly, another citation peak comes, on average, as late as ten years after the application of the patent.

While citation-weighted patent measures can reveal useful information about the innovation performance patterns of individual firms and industries, and the use of these measures has become more common (Walker, 1995), few studies have explored the issues relating to the construct validity of these measures. In the following sections I focus on two measurement validity issues: the correct citation lag (the number of years included after the patent's application to count citations; in figure 14.1 twelve

years were included), and the potential differences in citation patterns between qualitatively different innovations.

Sample study: biotechnology companies and radical innovation performance

In this section empirical data on biotechnology collaboration performance are examined. I test the effects of R&D collaborations on the radicality of innovation performance using patent data within a sample of 100 biotechnology companies. This study is used as an example of the use of patents for measuring innovation performance. Since the main focus of this contribution is on innovation performance measurement, I only briefly summarize the hypotheses of the sample study and the empirical results. In the following sections these sample data are further examined to test the validity of the patent-based radical innovation measures.

Main hypotheses

The main hypotheses of the sample study, and the empirical methods used, are summarized below. This sample examines R&D collaborations between small and large biotechnology companies. The study hypothesizes that collaboration can have negative effects on innovation performance. More specifically, collaboration with dissimilar partners – with older, larger or foreign partners – is hypothesized to have a negative effect on the radicality of the innovation output of the smaller partner. The effects of the number of collaborative partners and the complementarity of partner characteristics on the radicality of the smaller partner's innovation output are also examined. A more detailed description of the theoretical arguments leading to the hypotheses is available from the author, and similar arguments can also be found in other studies on R&D collaboration (see, for example, Katila, 1997, Katila and Mang, 2003, Lane and Lubatkin, 1998, and Shan, Walker and Kogut, 1994).

Three main hypotheses are examined.

- (1) *Hypothesis 1*. The number of collaborative partners has a curvilinear (inverted U) relationship with the radicality of the innovation output of the smaller partner. The radicality of innovations increases up to a point, but after this optimal point has been reached additional increases in the number of partners is negatively related with the radicality of innovation output.

- (2) *Hypothesis 2a.* The larger and older the R&D collaborative partner the less the radicality of the innovation output of the smaller partner.
Hypothesis 2b. Foreign R&D collaborative partners have a negative effect on the radicality of the innovation output of the smaller partner.
- (3) *Hypothesis 3.* Complementarity in the resources of the collaborative partners, such as complementary marketing and technological capabilities, has a positive effect on the radicality of the innovation output of the smaller partner.

Methods

To test these hypotheses, data on 100 biotechnology firms founded between 1980 and 1988 were gathered. The starting point of the study, 1980, was a significant milestone in the US biotechnology industry: the first genetically engineered organism was patented in that year. The sample includes biotechnology companies listed in *PaineWebber* and *Genguide* biotechnology-specific directories and for which sufficient data were available during the period of study. Only biotechnology firms concentrating on human therapeutics and *in vivo* diagnostics are included; this way the underlying technological setting and expertise requirements are relatively constant and the innovation output of the sample firms is comparable.

The dependent variable, the radicality of innovation output, is measured by citation-weighted patent counts. To distinguish between companies that produce incremental improvements and those that focus on more radical innovations, I weight the number of patents with the citations the patents receive during the six years after the application of the patent, or until the end of 1997. Self-citations are excluded from this data. Patent information was obtained from US Patent and Trademark Office (PTO) documents, and it includes yearly counts of the patents that the sample firms had applied for each year.

There are five independent variables in the study that measure different characteristics of the biotechnology firms' collaboration behaviour. The first independent variable, resource complementarity, measures the fit between the collaboration partners' resources. Resource complementarity is operationalized as an interaction between the larger partner's sales and marketing experience and the smaller partner's research capability. Research capability is measured as the cumulative number of the smaller partner's citation-weighted patents in three past years (years $t-3$ to $t-1$; see Henderson and Cockburn, 1994).

The remaining partner characteristics are partner experience (“partner age”), partner size (“partner sales”), a binary variable indicating a foreign partner, and the number of R&D collaborative partners. Due to the time-series nature of the data, a lagged-variable design is used: data for the independent variables are collected a year before the dependent variable values. The data for this study were collected from several biotechnology-specific databases and directories, annual reports and annual form 10-K reports in the United States of these companies, as well as from the PTO database. *PaineWebber* and *Genguide* directories, *Predicasts* and various news databases were the sources of the cooperation data. Data regarding the collaborative partners were drawn from the *Compustat* database, annual reports of the companies, and news articles in *Lexis Nexis*.

Poisson regression analysis was used to test the hypotheses. Poisson regression models have a number of attractive features for patent-based innovation measurement: they are suitable for integer data (counts of events), and they also cater for counts that are aggregated over time periods (McCullagh and Nelder, 1989). In this study, the dependent variable, “patents”, is a non-negative count of patents, and observations are combined to a time-series panel (yearly observations over a nine-year period for each firm). Poisson regression is therefore an appropriate method to use.

Results

The innovative output of the sample companies was highly diverse: on average, these companies applied for 1.3 patents yearly, although some had no patents in any year (eight companies), and one organization applied for 18 patents in a single year. The number of R&D collaborations examined in this study was 246, resulting in 894 yearly observations. Yearly data for the companies were collected in 1980–90, including patent citations until the end of 1997.

The results of the Poisson regression analysis are summarized in table 14.2. In all, the empirical results of this sample study show strong support for the above-mentioned hypotheses, and imply that unbalanced combinations between collaborative partners are likely to result in lower radicality of innovation output. Moreover, further analysis of the data demonstrates that patent citations play a significant role in the measurement: when raw patent counts instead of citation-weighted counts were used in testing the hypotheses, results become significantly weaker. More detailed results, as well as descriptive statistics, are available from the author.

Table 14.2: Results of the Poisson regression analysis predicting radicality of innovation

Variable	Proposed sign	Findings	Significance
H1: number of R&D partners	+/- (inverted U)	+/- (supported)	0.623*** (number) -0.007*** (number ²)
H2a: partner sales	-	- (supported)	-0.0001***
H2a: partner age	-	- (not significant)	-0.0006
H2b: foreign partner	-	- (supported)	-0.245***
H3: complementary resources	+	+ (supported)	0.00001***

Notes: Last column gives parameter estimates. *** $p < 0.001$ (two-tailed tests).

Analysis of patent-based measures

The above-discussed study on biotechnology patenting gives an example of the use of patents as a measure of the radicality of innovation output. However, the use of citation-weighted patent measures raises two measurement issues that are often overlooked. First, I contend that prior work using citation-weighted patent counts has customarily used rather short citation lags. For example, Dutta and Weiss (1997) use only five years of citation data. I contend that this lag may not be long enough to capture the differences in firm innovation performance. I know of no other work that has examined the effects of citation length on the validity of the empirical results.

Second, I ask whether the citation lags used in prior work are long enough to discriminate between incremental and radical innovations. If receiving more citations reflects the radicality of the patent, and the value of radical innovations is likely to be acknowledged relatively late after their introduction (Trajtenberg, 1990; Utterback, 1994), it is likely that the patents that eventually get the most citations are cited relatively late. Short lags would not therefore be likely to capture the majority of the citations that radical innovations receive, and thus would not accurately reflect their radicality. I hypothesize that the citation pattern of incremental innovation is skewed to the left, whereas radical innovations are likely to exhibit an opposite pattern.

To address these questions, two tests were conducted with the sample study data. The objective of these tests was to determine how sensitive the patent-based measure of innovation performance is to the citation lag used. In the first test I compared the average innovativeness of the sample companies in 1985–6 by using a citation-weighted patent count (see figure 14.2). A list of the

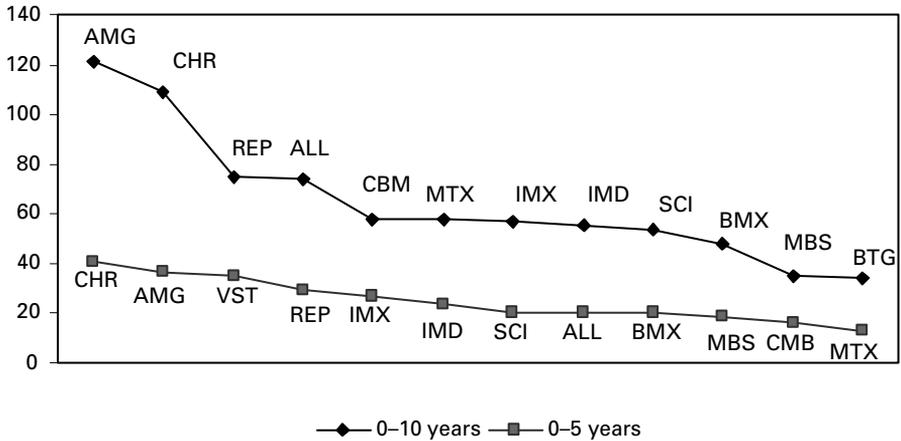


Figure 14.2: Most innovative biotechnology firms in 1985–6 by citation-weighted patent measures

Notes: Companies in “0–10 years” series: Amgen, Chiron, Repligen, Allelix, Creative Biomolecules, Matrix, Immunex, Immunomedics, Scios, Biomatrix, Molecular Biosystems and Bio-Technology General. Companies in “0–5 years” series: Chiron, Amgen, Vestar, Repligen, Immunex, Immunomedics, Scios, Allelix, Biomatrix, Molecular Biosystems, Creative Biomolecules and Matrix.

most innovative biotechnology companies in these years was first compiled using citations up to five years after the application of each patent. The second list was prepared using a period of ten years after application.

As shown in figure 14.2, a comparison of the twelve most innovative companies in both lists leads to the conclusion that the length of the citation period does have an effect on the innovation performance position of the company. Although many of the top companies change relatively little in their positions as the citation period is extended, some companies, such as Creative Biomolecules (fifth in “0–5 years” versus eleventh in “0–10 years”) and Vestar (fifteenth in “0–5 years” versus third in “0–10 years”) “become” fundamentally more or less innovative when the longer citation period is used. It is also interesting to note that the group containing the four most innovative companies becomes increasingly distinguished from the rest of the firms as the longer citation period is used. This brief test gives an indication that the citation-weighted measure is quite possibly sensitive to the citation period, and that five years may not be a long enough citation period in many cases.

In the second test I selected a subset of patents, applied in 1980–7, and calculated the average time period for these patents to receive all their citations. The sample was split in two parts based on the number of citations each company-year had received. The purpose of the test was to find out whether years that had received an above-average number of citations had

longer citation periods. Two benchmarks were used: receipt of 30 per cent and 80 per cent of the citations.

I first tested the proposition that radical innovations (most cited) are recognized and cited later than more incremental innovations. Indeed, it took significantly ($p = 0.019$) longer for those years that received an above-average number of citations to receive 80 per cent of the citations than it did for the less cited years. On average, it took 9.5 years for the less cited portfolios, versus 10.4 years for the more cited portfolios, to receive 80 per cent of the total citations. Thus, more heavily cited – i.e. technologically more radical – innovations tended to receive citations later than technologically incremental innovations. I also tested the same sample for the 30 per cent benchmark. Surprisingly, less cited patents received the first 30 per cent of citations later (3.7 years) than the more cited patents (2.5 years) ($p = 0.001$). In other words, the first citations for radical innovations were received early, but the majority of the citations were received relatively late.

In all, these two tests raise several empirical issues for performance researchers. Preliminary results on the sample of 100 biopharmaceutical companies in 1980–97 show that, in most cases, a citation lag of five years is not adequate to measure innovation performance reliably. In the case of biotechnology patents, for example, 80 per cent of the citations were not received until, on average, ten years after the patent application. Furthermore, I found that radical innovations have longer citation lags than more incremental innovations. The results indicate that using short citation periods may result in the exclusion of radical innovations from the sample. Consequently, researchers need both to ensure that the length of citations used is sufficiently long and to experiment with the effects of different citation lags on their results.

Conclusions

This chapter has three main aspects. First, from the theoretical standpoint, it presents an overview and structures the previous literature on innovation performance measurement and radical innovation. Four different categories of radicality are identified. This categorization is important for measuring radicality at the appropriate level of analysis that corresponds to the theoretical meaning of the construct. Patents and patent citations are suggested to be appropriate measures of innovation performance if the construct of interest is *technologically* radical innovation.

Second, this study contributes to subsequent work on innovation measurement that uses patents to measure the technological radicality of innovation performance. The chapter demonstrates the use of patent measures to analyse innovation performance. The main conclusion of the empirical analysis is that the length of the citation lag can dramatically change the picture of the innovativeness of firms. Moreover, the study shows how citation patterns of incremental and radical innovations differ: radical innovations tend to be cited later than incremental innovations. Consequently, short citation lags may not capture the value of radical innovations properly. Citation lags of ten years and longer are recommended for performance measurement.

Third, this work has managerial implications. The sample study on biotechnology collaboration presents evidence of the negative effects of collaboration on the radicality of innovation output. From the managerial perspective, these results complement those of a more recent study by Lane and Lubatkin (1998), who find that relative similarity in partner characteristics can enhance the smaller partner's innovative performance. This study therefore emphasizes the need for small organizations to select their partners carefully in order to obtain maximum radical innovation performance. The measurement issues discussed in this study can help in further clarifying the sources and measurement of radical innovation performance for these companies.

References

- Achilladelis, B., Schwarzkopf, A., and Cenes, M. (1990). The dynamics of technological innovation: the case of the chemical industry. *Research Policy*, 19, 1–34.
- Ahuja, G. (1996). Collaboration and innovation: a longitudinal study of interfirm linkages and firm patenting performance in the global advanced materials industry. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.
- Albert, M., Avery, D., Narin, F., and McAllister, P. (1991). Direct validation of citation counts as indicators of industrial patents. *Research Policy*, 20, 251–9.
- Anderson, P., and Tushman, M. (1990). Technological discontinuities and dominant design: a cyclical model of technological change. *Administrative Science Quarterly*, 35, 604–33.
- Arora, A., and Gambardella, A. (1994). The changing technology of technological change: general and abstract knowledge and the division of innovative labor. *Research Policy*, 23(5), 523–32.
- Carpenter, M., Narin, F., and Wolf, P. (1981). Citation rates and technologically important patents. *World Patent Information*, 3, 160–3.
- Christensen, C. M., and Rosenbloom, R. S. (1995). Explaining the attacker's advantage: technological paradigms, organizational dynamics, and the value network. *Research Policy*, 24(2), 233–57.

- Dutta, S., and Weiss, A. (1997). The relationships between a firm's level of technological innovativeness and its pattern of partnership agreements. *Management Science*, 43(3), 343–56.
- Green, S., Gavin, M., and Aiman-Smith, L. (1995). Assessing a multidimensional measure of radical technological innovation. *IEEE Transactions on Engineering Management*, 42(3), 203–14.
- Griliches, Z. (1984). Introduction. In Griliches, Z. (ed.), *R&D, Patents and Productivity*, 1–20. Chicago: University of Chicago Press.
- Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry. *Rand Journal of Economics*, 24(2), 248–70.
- Henderson, R., and Clark, K. (1990). Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 35, 9–30.
- Henderson, R., and Cockburn, I. (1994). Measuring competence? Exploring firm effects in pharmaceutical research. *Strategic Management Journal*, 15 (Special Issue), 63–84.
- Jaffe, A.B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3), 577–98.
- Katila, R. (1997). Technology strategies for growth and innovation: a study of biotechnology ventures. In Reynolds, P. D., Bygrave, W. D., Gartner, W. B., Carter, N. M., Mason, C. M., Davidsson, P., and McDougall, P. P. (eds.), *Frontiers of Entrepreneurship Research – 1997*, 405–18. Wellesley, MA: Babson College.
- (2002). New product search over time. *Academy of Management Journal*, 45, 995–1010.
- Katila, R., and Mang, P. (2003). Exploiting technological opportunities: the timing of collaborations. *Research Policy*, 32(2), 317–32.
- Kortum, S., and Lerner, J. (1999). What is behind the recent surge in patenting? *Research Policy*, 28(1), 1–22.
- Lane, P., and Lubatkin, M. (1998). Relative absorptive capacity and interorganizational learning. *Strategic Management Journal*, 19, 461–77.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Mitchell, W. (1989). Whether and when? Probability and timing of incumbents' entry into emerging industrial subfields. *Administrative Science Quarterly*, 34, 208–30.
- Pavitt, K. (1988). Uses and abuses of patent statistics. In Van Reenen, A. (ed.), *Handbook of Quantitative Studies of Science and Technology*, 509–36. Amsterdam: Elsevier Science.
- Rosenbloom, R., and Christensen, C. (1994) Technological discontinuities, organizational capabilities and strategic commitments. *Industrial and Corporate Change*, 3(3), 655–85.
- Shan, W., Walker, G., and Kogut, B. (1994). Interfirm cooperation and startup innovation in the biotechnology industry. *Strategic Management Journal*, 15, 387–94.
- Trajtenberg, M. (1990). *Economic Analysis of Product Innovation: The Case of CV Scanners*. Boston: Harvard University Press.
- Utterback, J. (1994). *Mastering the Dynamics of Innovation*. Boston: Harvard University Press.
- Walker, R. (1995). *Patents as Scientific and Technical Literature*. Lanham, MD: Scarecrow Press.

Graham Clark and Antony Powell

Introduction

This chapter introduces a method for collecting performance data in context. The method, termed Context-based Measurement (CbM), preserves a rich description of the circumstances of a measure. This contextual information is often overlooked, but essential in preserving the meaning of measurement information.

The principle of CbM can be explained using the metaphor of the Global Positioning System (GPS). In aircraft navigation, for example, a GPS receiver derives the speed and direction of an aircraft by recording over time the coordinates of longitude, latitude and altitude. This contextual data is then used to derive the distance to the destination, communicate progress to date and calculate the optimum flight path.

Now imagine if we had an equivalent of the Global Positioning System for the measurement of organizations. The coordinates of longitude, latitude and altitude might be replaced by ones of resource used, process performed and product produced. If we designed a GPS for an organization, we could measure its *performance* (e.g. cost or quality) in the context of the resource used, the process performed and the product delivered. Such an approach could help us specify our targets, communicate our performance and signal our strategy. This is context-based measurement.

The remainder of this chapter explains the method of CbM in more detail using a case study in the automotive industry. The reader should note, however, that the principles are not specific to an industry or discipline. We start by explaining how structures are essential to understanding in any form of measurement.

Using structures for measuring

All measures have a context, and without adequate context our measures are meaningless. This knowledge of context gives meaning but all too often organizations depend on measurements in contexts that are assumed and inherited. For example, measures of *profit*, *balance* and *return on investment* rely on generally understood accounting definitions and standard units.

These *measurement frameworks* allow us economies in the amount of description required to give specific meaning. If we want to know something more specifically (e.g. quarterly profit on component X) then we need to describe and collect our measurements with more specific context. However, whilst we have formalized structures for organizational accounting, we argue that the same level of descriptive rigour is not given to other structures, such as work breakdowns, resource breakdowns, quality, etc. That is, we need to make sure our measurements have sufficient context to be understood.

The use of measurement is a fundamental element of engineering disciplines (De Marco, 1982). However, because engineering is a human-led, design-driven activity its measurement is inherently difficult. This is because the resources, processes and work products change over the course of the project, and the product evolves via lifecycle incarnations from concept to end-solution. This might not be a problem when the activity is well practised and the context understood, but when the activity is being performed for the first time the absence of previous knowledge can threaten the outcome of the project. The response has been to introduce more disciplined approaching to controlling activity. At the business level this includes enterprise resource management (ERM) applications for asset and cost accounting. At the engineering level are structures of product assemblies, services and budgets, with a primary structure being the work breakdown structure, or WBS (Tausworthe, 1980).

A WBS is a product-oriented hierarchy of all the work to be carried out in order to complete the contractual goals of a project. The structure progressively decomposes the design of a whole product into sub-products, and then down to work packages that constitute a controllable quota of resources required to complete a specific task (see figure 15.1). The WBS provides the context against which budgets are planned and spend is compared. The correct degree of refinement in the definition of work packages in the WBS pays dividends over the life of a project, as it will ensure the correct level of management control of the activity.

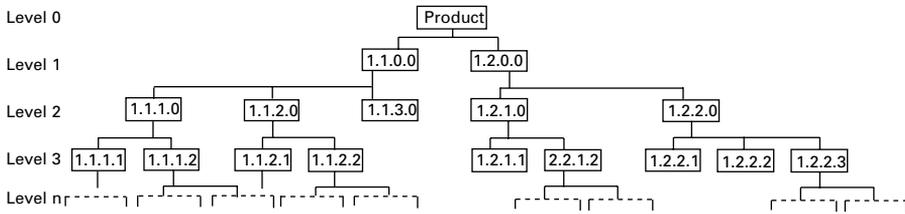


Figure 15.1: General breakdown structure numbering convention

The introduction of earned value measurement (EVM) has improved project reporting by passing performance measurement targets, protocols and scales down through the WBS decomposition to the processes being managed (Fleming and Koppelman, 2000; Turner, 1993). Nonetheless, it is these specific characteristics that require detailed measurement in order to provide management with rich enough data for incorporating change, re-estimating and replanning.

While the WBS serves well in defining all the work to be achieved and acts as a vital communications tool for projects (Kerzner, 1998), a poor structure can obscure vital management information during the lifetime of a project. The method of naming each work package in the WBS is less well defined and is usually limited to a verb and a noun – e.g. *design product*, *review product* – or sometimes just a noun. This remains an issue within the project management profession, as it seems incongruous that such an important project asset as work package naming is reduced to such brevity (Berg and Colenso, 2000). Rook observed that “many of the most significant activities” in a project, such as configuration, assurance and liaison, do not conform to the functional structure of a WBS, and as a result it is difficult to collect data on across the whole or parts of the product during its development (Rook, 1991).

The WBS illustrated in figure 15.2 collects effort data across a mixture of product and process descriptions. By aggregating the branches, it allows the total cost of specific projects to be established, but it is extremely hard to find the cost of the constituent development activities. The way a WBS is organized, therefore, has a major influence on the way data is collected, and, critically, on the types of questions that can be asked later about the process.

When projects benefit from previous experience a coarse level of measurement is usually sufficient to manage the project successfully. Conversely, when the work is novel or more research-and-development-oriented, then gauging the progress and value of the work accomplished is more challenging,

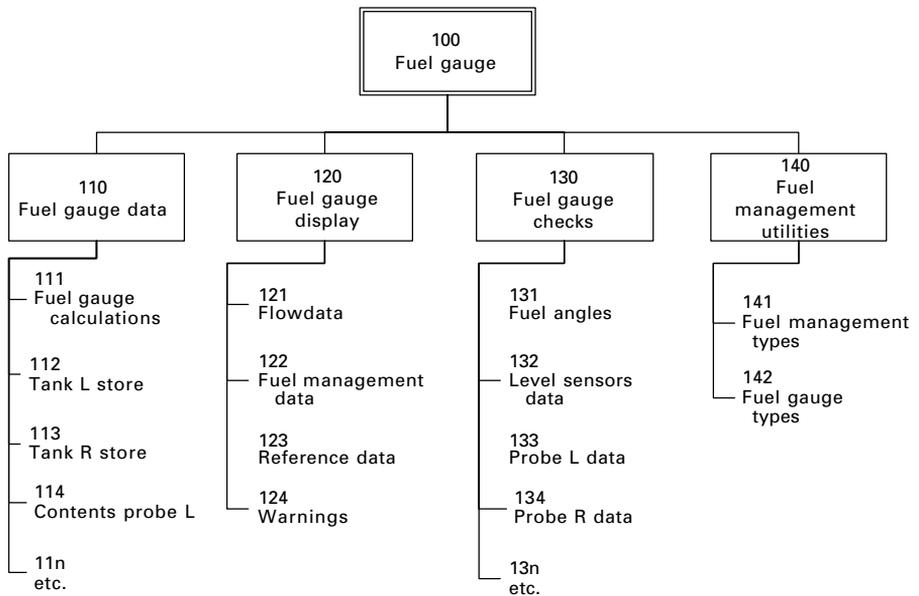


Figure 15.2: Excerpt from fuel gauge breakdown structure

since there is a greater likelihood of changes to the product as problems present themselves (Powell, 2001). When this is the case, management often want closer control and detailed measures to track the work and watch for early deviations from planned activities. This can increase the size of the structure dramatically, with consequences for administrative overheads, particularly when the nature of the project is dynamic or has a high degree of associated uncertainty. Consequently, a tension develops between the desire for management insight and the collection overhead of detailed measures across a large organization. This is partly because of the naming conventions in the WBS, but mainly it is due to the static nature of its tree-like structure, which represents *structural levels* in a mono-hierarchy that doesn't reflect the *descriptive levels* in organizations (Ahl and Allen, 1996). In other words, the hierarchical WBS is akin to a GPS that captures only longitude *or* latitude, but not both.

Meaningful performance measurement relies on structured measurement frameworks to connect corporate goals to operations (Neely *et al.*, 1997; Medori and Steeple, 2000). We conclude that measuring activities that are locked into static WBS is difficult and a limiting factor in reporting performance measurement in project management. Instead, we need a measurement approach that reflects the business and technical context of a measure.

Case study

The need to measure in context can be illustrated by a case study from the automotive industry. This study examines the improvement of measurement on a project to develop a fuel system for a high-performance car. The car had two fuel tanks that were linked to the engine and each other for balance and cooling, automatic control of the valves, and instrumentation in the car's cockpit for the driver. The monitoring and active management of fuel meant that it was heavily dependent on software – a critical aspect of the safety of the vehicle.

The management and technical staff wanted accurate, meaningful data to support local estimating, but were experiencing problems trying to record what activities they had carried out. The effort collection bookings for staff were to be expressed in the work package codes available within the WBS (see figure 15.2), where the product is decomposed into subsystems, with a coding system to distinguish and classify the components. For this part of the project, the WBS listed over 300 booking numbers in the cost booking system, and, because the structure was so large, very few staff were able to locate a meaningful activity against which to record their effort. It transpired that, in this part of the project, 73 per cent of all booked hours were recorded against just twelve WBS numbers. These were high-level activities that covered large areas of work and were unsuitable for detailed estimating.

While the project needed some regulation over what technical staff could book to, the staff also needed to be able to identify the tasks they had performed in an efficient structure that was meaningful to them. They also needed to be able to express the work that had actually been done, rather than what the WBS dictated through its structure. Although this WBS had some activity-oriented tasks, because the levels of process were not always the same as those of the product there was a great deal of repetition of terms throughout the structure. Segregating process from product would be difficult, especially as the work involved a number of development stages and design representations. These inherent weaknesses and the lack of rigour in work breakdown structures meant that the data collected were both limited and subject to interpretation.

In summary, the project managers wanted accurate, meaningful data to support the planning, estimating and monitoring of achievement of the project; but hardly anyone in the teams was able to record satisfactorily what activities they had performed against the WBS effort booking structure. They

were unable to express the work that had actually been done, rather than what was expressed through the WBS.

A review of the content of the WBS and the project measurement method was required in order to provide a performance measuring system that staff could use with ease but that delivered fine-grained measures for decision making.

Context-based measurement

CbM is a language-based measurement approach that preserves rich descriptions of the circumstances in which a measure is captured. Like any language, CbM consists of a lexicon and a grammar.

The *lexicon* is essentially a constrained vocabulary of terms (verbs and nouns) that are used to describe a process. Here we distinguish a “term” from a natural-language “word” by assigning it a single meaning in the lexicon. The basic language in our case study consists of four “models” common to most organizations: processes, products, product representations and stages.

- (1) *Processes* are the generative, evaluative and supportive verbs that describe what we do. For example: define, design, produce, review, manage, support, etc.
- (2) *Products* are the configurable physical or functional components or systems that are produced. For example: car, engine, fuel system, starting system, etc.
- (3) *Product representations* are the intermediate outputs (work products) of a process. For example: designs, models, software code, estimates, etc.
- (4) *Stages* are time-based (milestone or cost) breakdowns of logically sequenced partitions of work. For example: project A, delivery 2, etc.

Each list effectively serves as a data dictionary that, when carefully managed, can be used to form a consistent description of process.

The *grammar* connects the lexical terms together by composing a permutation across the four dimensions shown in table 15.1. The resulting description essentially gives us coordinates in the “process space” (akin to longitude, latitude and altitude) against which the various properties of a process (e.g. effort, size, time) can then be measured.

The benefit of CbM arises from its ability to support the *aggregation* of information and the *integration* of process and product models. The dimensions of *stage* and *product* contain an inherent hierarchy through which information can be *aggregated*. For example, we can view product

Table 15.1: Example CbM lexicon

		CbM for twin-tank fuel gauge cost collection issue 1.3	
Process	Product	Product representation	Stage
	TTFGS		
Process independent	(twin-tank fuel gauging system)		
Analyse	Product independent	Representation independent	Phase independent
Approve	FUEL GAUGE	Avionic system test instructions	Definition
Assure	FUEL GAUGE DATA	Configuration documents	Detailed design
Code	FUEL GAUGE CALCULATIONS	Design	Estimate
Collate	TANK L STORE	Detailed design document	Feasibility
Configure	TANK R STORE	Equipment specification documents	Handover
Define	CONTENTS PROBE L	Instrument layout	Integration testing
Estimate	CONTENTS PROBE R	Instrument lighting	Pre-feasibility
Integrate	DATA HISTORY	Interface control	Preliminary design
Liaise	TOTAL FUEL LEVEL DATA	Load analysis document	Requirements capture
Maintain	DATA TABLES	Operations environment document	Road test
Manage	TOTAL CONTENTS DATA	Preliminary design document	Software code
	FUEL GAUGE DISPLAY	Procedures	System requirements
Meet	FLOW DATA	Quality documents	requirements
Operate	FUEL MANAGEMENT DATA	Requirements document	Testing
Partition	REFERENCE DATA	RIG requirements document	Vehicle integration
Plan	WARNINGS	Safety clearance document	
Produce	FUEL GAUGE CHECKS	Source code units	
Rationalize	FUEL ANGLES	Standards	
Report	LEVEL SENSORS DATA	Subsystem	

Review	133	PROBE L DATA	Symbology
Specify	134	PROBE R DATA	System definition document
Support	135	TANK STATUS	System drawings
Test	136	SERVICE CHECK WARNING	Tank simulation
Write	137	PRESSURE VENT VALVE	Test
	138	TANK INTERCONNECTION VALVE	Test acceptance
	139	TRANSFER ISOLATION VALVE	Test equipment
	13A	VENT BYPASS VALVE	Test report
	13B	VALVE MANAGER	Testing document
	140	FUEL MANAGEMENT UTILITIES	Vendor
	141	FUEL MANAGEMENT TYPES	
	142	FUEL GAUGE TYPES	

Table 1 Example CbM Lexicon

effort at the levels of function (e.g. FUEL GAUGE) or subsystem (e.g. TANK L STORE). Similarly, we can view the stage effort at the levels of task, delivery, project or programme. The dimensions of *process* and *product representation* describe operational activities as performed by engineers, such as “Analyse source code” or “Approve design”. We can, therefore, view the effort of each process *stage*, such as “System requirements” and “Vehicle integration”. Alternatively, we can view the relative spend on each process – for instance, evaluating the time spent to produce and review a particular product through all its *product representations*.

CbM integrates the models of process and product, as each sentence captures both the work performed and the process through which it was performed. The “language”, therefore, provides a description of process that is more semantically rich than its constituent elements. This generic WBS can, if carefully managed, form a consistent and rigorous model of process as a basis for description and measurement.

The process of acquiring the descriptive data necessary for a CbM statement involved harvesting terms for our controlled vocabulary and building relationships between them. In order to produce a new performance measurement system for this project, we stepped through the overall CbM process for the definition (solid lines) and maintenance (broken lines) as presented in figure 15.3. We assumed that some degree of project organization was in place with adequate project structures and dedicated roles for management and control. The first step in building descriptions of activities for staff to book to was to acquire project terms from existing project procedures, standards, manuals, guidelines and other documents and organizing them into a controlled vocabulary for the project (steps 1–4). Subsequent presentation and collection (steps 5–6) followed as a plan–do–act–check cycle (Deming, 1986) in which new requirements for measurement could be incorporated and the project vocabulary maintained.

We now explain the process with a six-step model, using the case study.

Determining dimensions and acquiring terms

This involved harvesting descriptive terms from existing organizational information. In this case we knew we had:

- (1) process-based activities;
- (2) product components – from the product breakdown structure;
- (3) design representations of the product; and
- (4) phases or stages of development.

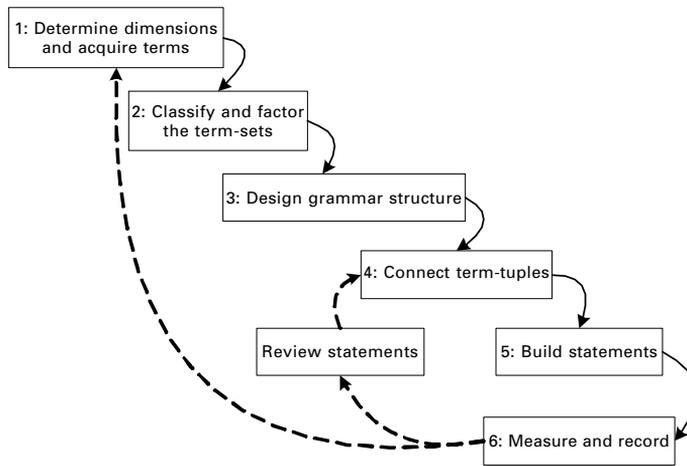


Figure 15.3: CbM construction

These different types of project information were classified into CbM dimensions: “process”, “product”, “product representation” and “stage”. To gain more detail, we collected other descriptive terms from local documents and interviews with staff.

Classifying and factoring the term-sets

We arranged the lists of words from our controlled vocabulary under the four dimensions. We discarded duplicate terms and ensured that contributors and users agreed definitions (see table 15.1, taken from the project spreadsheet). We organized the “process”, “product representation” and “stage” dimensions into alphabetic lists, but kept the “product” list in its original breakdown structure form, as in figure 15.2, in order to preserve the project WBS coding policy. With this approach, the extensive WBS for this part of the project was reduced to a single sheet of paper.

Designing grammar structure

Next we designed a simple grammar structure to act as a guide to the ordering of the dimensions:

<PROCESS> the <PRODUCT> <REPRESENTATION> in the <STAGE> stage for <Hours> hours.

We needed a statement that had enough information about the circumstances of the project, when the measure was taken, for use in queries and estimates.

Connecting term-tuples

The terms contained in chevrons above had associations made with the other terms in the other lists by the team members, as in figure 15.4. In this way a potential set of over 300,000 associations was reduced to a constrained vocabulary.

Building statements

We had to enter the effort spent on tasks during a working week. In figure 15.5, statements have been entered into a web-based application called CONTEXT (YorkMetrics, 2005). It constructs CbM statements using interchangeable panels. The “declare” panel has pull-down menus to describe activities and record a measurement against them. The “measure” panel, here an effort booking panel with decimal hours but interchangeable with other scales, presents the user’s current statement prior to committal to a database. Completed statements and measures are visible in the “observe” panel.

In the panels in figure 15.5, a user has reported an effort measure. He or she initially selected the “analyse” term, which constrained the next list to include the “fuel gauge calculations” term, and he/she is now able to select from what is available in the “representations” list before selecting the stage and declaring his/her effort on this activity for the week. Completed statements and measures are visible in the lower panel.

Measuring and recording

The measures are then recorded to a database and coordinated by team leaders, who check and underwrite the booking records made. In the same way, CONTEXT could be used to record other measures for the project using other grammars and lexicons, including risk, defects, maturity, etc. Using this six-step model, the measuring system could be tailored for other projects with their own project vocabulary, and cross-project queries could be undertaken.

Process	Product	Product representation	Stage
Process independent	TFFGS (twin-tank fuel gauging system)	Representation independent	Phase independent
Test	000 FUEL GAUGE	Architecture	Pre-feasibility
Analyse	110 FUEL GAUGE DATA	Configuration documents	Feasibility
Approve	111 FUEL GAUGE CALCULATIONS	Design	Definition
Assure	112 TANK L STORE	Detailed design document	Estimate
Code	113 TANK R STORE	Equipment specification documents	Requirements capture
Collate	114 CONTENTS PROBE L	Instrument layout	System requirements
Configure	115 CONTENTS PROBE R	Instrument lighting	Preliminary design
Define	116 DATA HISTORY	Interface control	Detailed design
Estimate	117 TOTAL FUEL LEVEL DATA	Load analysis document	Software code
Integrate	118 DATA TABLES	Operations environment document	Testing
Liaise	119 TOTAL CONTENTS DATA	Preliminary design document	Integration testing
Maintain	120 FUEL GAUGE DISPLAY	Procedures	Vehicle integration
Manage	121 FLOW DATA	Quality documents	Road test
Meet	122 FUEL MANAGEMENT DATA	Requirements document	Handover
Operate	123 REFERENCE DATA	RIG requirements document	
Partition	124 WARNINGS	Safety clearance document	
Plan	130 FUEL GAUGE CHECKS	Source code units	
Produce	131 FUEL ANGLES	Standards	
Rationalize	132 LEVEL SENSORS DATA	Subsystem	
Report	133 PROBE L DATA	Symbology	
Review	134 PROBE R DATA	System definition document	
Specify	135 TANK STATUS	System drawings	
Support	136 SERVICE CHECK WARNING	Tank simulation	
Write	137 PRESSURE VENT VALVE	Test	
	138 TANK INTERCONNECTION VALVE	Test acceptance	
	139 TRANSFER ISOLATION VALVE	Test equipment	
	13A VENT BYPASS VALVE	Test report	
	13B VALVE MANAGER	Testing document	
	140 FUEL MANAGEMENT UTILITIES	Vendor	
	141 FUEL MANAGEMENT TYPES		
	142 FUEL GAUGE TYPES		

Figure 15.4: Pairing the terms in adjacent dimensions

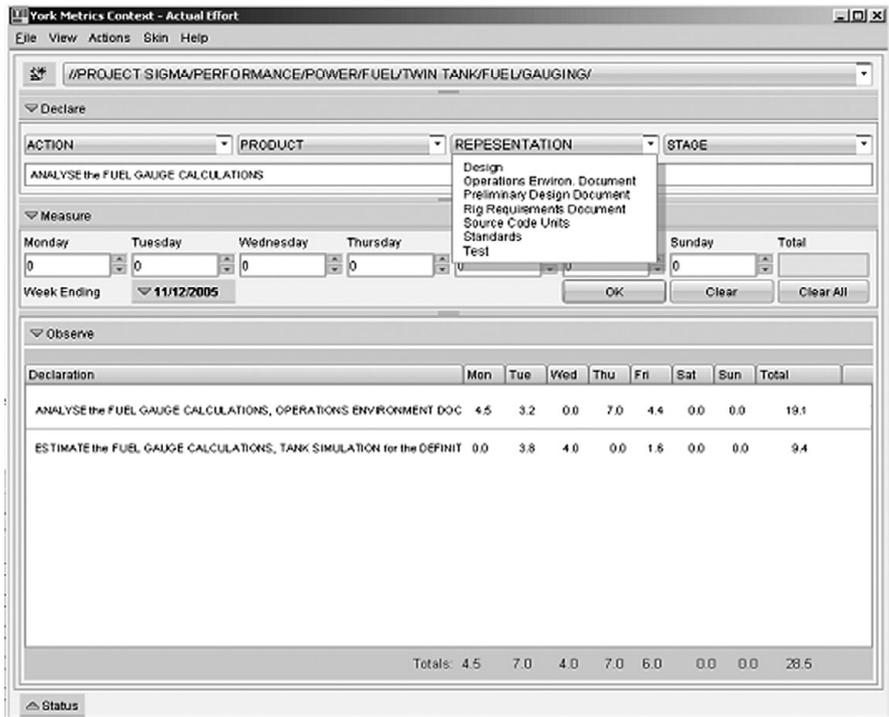


Figure 15.5: Software application using CbM to measure activities (www.yorkmetrics.com)

The benefit of having an activity record of this richness is that a mix of evidence from project development phase, process and product information is presentable in a meaningful manner, using local project terms to describe a measurement in its context. The measures are capable of being aggregated within the structure to satisfy corporate accounting, but they are also available in discrete dimensions for specific questions, and could be integrated into database queries to yield insights into project behaviour.

Results

Querying the large amounts of data was also significantly eased with CONTEXT. Figure 15.6 presents a graph of selected efforts expressed through the dimensions of “stage” and “process” from our case study. It is apparent from the figure that process activities were being performed out of their implied stages, and this was the basis for finer-grained queries. The activity of “coding”, naturally, took place in the “coding stage”, but it was

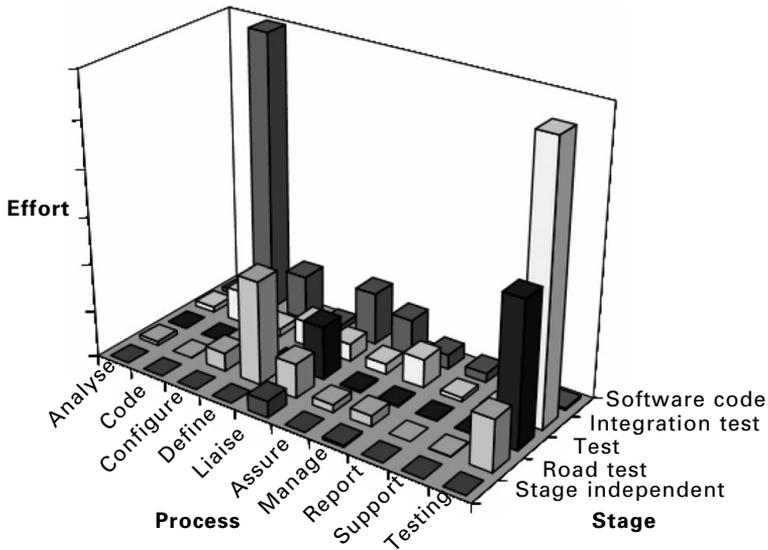


Figure 15.6: Effort in activities across the process and stage dimensions

also evident in the “integration test stage”. The constrained language was used with wildcards (“*”) to construct context-based queries for the project vocabulary and applied to the measurement database. For example:

<ANALYSE> the <FUEL GAUGE CALCULATIONS> <TEST DOCUMENTATION> in the <DEFINITION> Stage.

This query is very specific and reports a very precise value with a well-defined context. It would present those hours expended on analysing the test documentation for the “fuel gauge calculations”, during the definition “stage”.

By using a wildcard value we can broaden the query, as in:

<REVIEW> the <LEVELS SENSOR DATA> <*> in the <*> Stage.

This would report how much effort had been expended reviewing all materials associated with the “levels sensors data”, across all “stages”. If this is compared with the effort associated with other process activities, it might provide some insight as to where process improvements could be made. Equally, it might provide information that could be used as a basis for benchmarking this project’s processes with others. Other wildcards reveal if any unplanned *coding* activity had occurred in the latter stages of the project:

<CODE> the <*> <*> in the <INTEGRATION TEST> Stage.

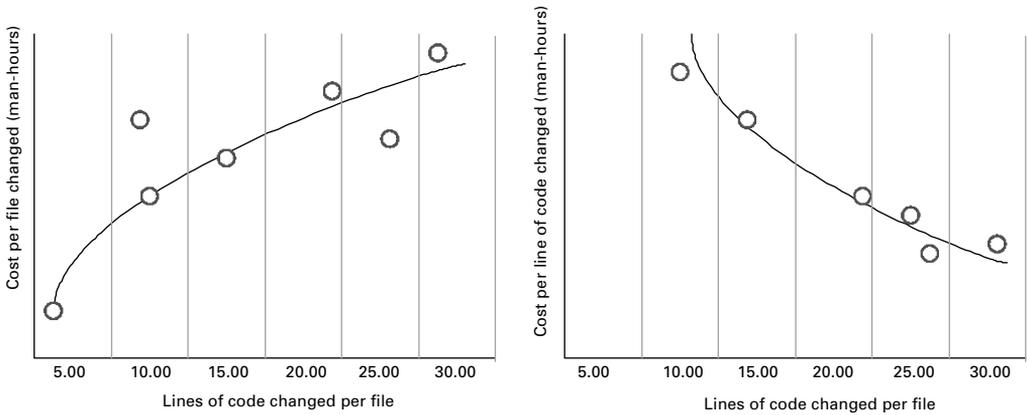


Figure 15.7: Cost of change to software components

We can also see what other activities occurred in the “coding stage”, including “configuration”, “assurance” and “liaison”, giving rise to a number of questions.

- Should we expect that much “coding” going on in the “integration test stage”?
- How much did we spend on “liaison” across the “life cycle” overall?
- Given the cost, can we rationalize the “test” activities?

However, two fundamental questions emerged in the case study.

- (1) What was the level of change?
- (2) What was the cost of handling it?

The team wanted to gain an insight into the amount of change that was qualitatively evident from using a change policy that queued up changes as they arose. By coupling high-resolution process measurement (here, effort) to product measurement (here, code size and code change measures), we saw that the relative cost of administering change to software was related not so much to the size of the amount of software changed but more to the complexity and degree of coupling that the software components had with other software components (Kamel, 1987). Highly complex and tightly coupled components were found to be the most expensive to change, and improvements in design to reduce these properties were taken as the opportunities arose. The size of the specific internal change to the component became a secondary cost driver.

Figure 15.7 shows how the cost of changing lines of source code was not linear and became proportionally amortized as more code was changed and, counter-intuitively, as the *cost per line of code changed* reduced with size,

thereby making large changes to a software component more economic than small ones. As the cost of performing change on software components became visible through the CbM data, the implication of a dominant fixed overhead cost in opening a software component for modification became the primary cost driver per object.

As a result of using these measures, the software team now knew when sufficient requests for change had been accumulated for a particular software component to make it economic to open it to change, add or delete lines of code, and test and then reintegrate it (Jørgensen, 1995). Tolerances of the estimates were established for changes to software, and the technical staff gradually reduced the estimate variance for changes by up to 38 per cent. Through knowing the cost of changing a software component and the criticality of the change, the team was able to estimate and organize the flow of change such that better decisions were made regarding the choice of software changes that the budget would support. The customer and the designers could now be asked to order the critical changes, such that change was both effective and efficient.

Conclusions

We have shown that project management information retrieval can be difficult because of the increased number of interrelationships between development activities and sub-products. Traditional WBS procedures provide only for the simple aggregation of data, resulting in many observations being harder to obtain, or passing unnoticed. We have examined the problems associated with orthodox project structures and found fine-grained measures in large, inflexible structures. For all but the most basic projects they become unmanageable, and the benefits of accurate measurement are soon outweighed by increases in the overhead of collection and administration. In demonstrating the value of providing a context for a performance measurement, we have highlighted its absence from traditional business structures.

Our performance measurement case study was a project experiencing a high degree of change in its development stages. In it, we applied a refined classification model using a controlled project vocabulary (Vickery, 1960; Doyle, 2003) to existing structures in which language helps projects to define themselves (Finkelstein and Leaning, 1984). We have shown how to design a “GPS” that couples technical and management activities under operational conditions in a simple, accurate and repeatable way (Pattee, 1985). Using

this, the effort expenditure on change activities was analysed and decisions were made to shift the locus of control for managing software product changes from the customer and management to the development team (Roche and Jackson, 1994).

This approach has achieved three principal goals. First, it enabled more fine-grained measures to be made without an increase in the cost of collection. Second, the CbM model was aligned with the process, product and project structure, allowing it to capture the precise context of each measure and present data at any aggregated level in the respective hierarchies. Third, by facilitating the integration of resource, process and product metrics it was able to separate cause and effect concerns. The benefits of this approach mean that many of our observations would have been much more difficult to obtain, or might even have been obscured, by conventional measurement approaches. A commercial toolset has extended the method to other applications and has provided a basis for future benchmarking.

References

- Ahl, V., and Allen, T. F. H. (1996). *Hierarchy Theory: A Vision, Vocabulary and Epistemology*. New York: Columbia University Press.
- Berg, C., and Colenso, K. (2000). "Work breakdown structure practice standard project: WBS vs. activities. In *Project Management Standards*, 69–71. Newtown Square, PA: Project Management Institute.
- DeMarco, T. (1982). *Controlling Software Projects: Management, Measurement and Estimation*. New York: Yourdon Press.
- Deming, W. E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Press.
- Doyle, R. (2003). The classification of content management systems. *Gilbane Report*, 11(2), 2–13.
- Fleming, Q. W., and Koppelman, J. M. (2000). *Earned Value Project Management*. Newtown Square, PA: Project Management Institute.
- Finkelstein, L., and Leaning, M. (1984). A review of the fundamental concepts of measurement. *Measurement*, 2(1), 25–34.
- Jørgensen, M. (1995). Experience with the accuracy of software maintenance task effort prediction models. *IEEE Transactions on Software Engineering*, 21(8), 674–81.
- Kamel, R. F. (1987). Effect of modularity on system evolution. *IEEE Software*, 4(1), 48–54.
- Kerzner, H. (1998). *In Search of Excellence in Project Management: Successful Practices in High Performance Organizations*. New York: Van Nostrand Reinhold.
- Medori, D., and Steple, D. (2000). A framework for auditing and enhancing performance measurement systems. *International Journal of Operations and Production Management*, 20(5), 520–33.

- Neely, A.D., Richards, A.H., Mills, J.F., Platts, K.W., and Bourne, M.C.S. (1997). Designing performance measures: a structured approach. *International Journal of Operations and Production Management*, 17(11), 1131–52.
- Pattee, H.H. (1985). Universal principles of measurement and language function in evolving systems. In Casti, J., and Karlqvist, A. (eds.), *Complexity, Language, and Life: Mathematical Approaches*, 268–81. Heidelberg: Springer-Verlag.
- Powell, A. L. (2001) Right on Time, Modelling and Managing Time-constrained Software Development, DPhil Thesis, Department of Computer Science, University of York, 2001.
- Roche, J., and Jackson, M. (1994). Software measurement methods: recipes for success. *Information and Software Technology*, 36(3), 173–89.
- Rook, P. (1991). *Software Engineer's Reference Book*. Oxford: Butterworth-Heinemann.
- Tausworthe, R. C. (1980). The work breakdown structure in software project management. *Journal of Systems and Software*, 181, 184–6.
- Turner, J.R. (1993). *The Handbook of Project-based Management*. London: McGraw-Hill.
- Vickery, B.C. (1960). *Faceted Classification: A Guide to the Construction and Use of Special Schemes*. London: Aslib.
- YorkMetrics (2005). CONTEXT. Performance Measurement software by YorkMetrics Ltd. www.yorkmetrics.com.

PART IV

Performance measurement in public services

Part III of this book explored some of the generic practicalities and challenges of performance measurement. Since the first edition of this book was published there has been an explosion of interest in the measurement of performance in public services, and so it is to this topic that we now turn our attention. This part consists of four chapters, three of which reflect on the challenges of measurement in different public service settings – education, the police force and health. The fourth contribution explores issues of perversity in public service performance measurement.

The opening chapter, by Andrew Brown, explores performance management in English primary schools. Brown identifies five reasons why performance might be measured: (i) measuring for the purposes of focusing attention; (ii) measuring to improve pupils' academic progress; (iii) measuring for professional development purposes; (iv) measuring for strategy implementation purposes; and (v) measuring to implement performance-related pay. Using a combination of desk and field research he explores the theoretical and practical challenges of each of the reasons for measurement, pointing out the inherent tensions and conflicts between them. Brown's chapter is important because it elucidates many of the reasons why performance measurement and management are so problematic in the English education system.

The second contribution, from Paul Collier, explores a different public service regime: the police force. Collier begins by tracing the evolution of performance measurement in the police. He then uses Foucault's framework to explore issues of sovereignty, discipline, governmentality and power in relation to police performance measurement, arguing that many of the changes that can be observed are a function of shifting balances between sovereignty, discipline and governmentality.

The third chapter explores measurement in the health care system. This chapter, by Rowena Jacobs, Peter Smith and Maria Goddard, discusses the challenges of constructing composite indicators to measure health care performance. Jacobs, Smith and Goddard propose an eight-step process for constructing composite indicators: (i) choosing the entities to be assessed; (ii) choosing the organizational objectives to be encompassed in the composite; (iii) choosing the indicators to be included in the composite; (iv) transforming measured performance on individual components; (v) combining the components, using some sort of weighting; (vi) adjusting for environmental or other uncontrollable influences on performance; (vii) adjusting for variations in expenditure (if a measure of efficiency is required); and (viii) conducting a sensitivity analysis to test the robustness of the outcome of the composite to the various methodological choices. They end their chapter by presenting a selection of composite indicators used in various countries, highlighting the strengths and weaknesses of these.

The final chapter in part IV, by Mike Pidd, explores perversity in performance measurement. Pidd reviews the challenges of performance measurement in the public sector, drawing on Hofstede's framework of control. Central to Pidd's argument is the assertion that, too often, a simplistic cybernetic control model underpins performance measurement systems. He argues that this model, while appropriate in some circumstances, can result in dysfunctional consequences in situations of uncertainty and ambiguity. As a result, he calls for a more thoughtful discussion of the appropriate forms of control in the public services.

Measuring the performance of England's primary school teachers: purposes, theories, problems and tensions

Andrew Brown

Introduction: background and historical context

The past decade has witnessed a revolution in how public services are managed around the world. Although a variety of slogans and labels, such as “new public management” (NPM; see, for example, Hood, 1991, Dunleavy and Hood, 1994, and Lane, 2000) and “reinventing government” (Osborne and Gaebler, 1992), have been used in different countries to describe their various public sector management reforms, many of the characteristics are similar. Kettl (2000), for instance, identifies six common core ideas: the search for greater productivity; more public reliance on private markets; a stronger orientation towards service; more decentralization from national to “sub-national” governments; increased capacity to devise and track public policy; and tactics to enhance accountability for results.

With regard to national education systems, Moos and Dempster (1998) argue that different countries and different segments of the public sector in those countries have introduced their own forms of new public management, with particular versions being influenced by a combination of historical and cultural conditions, political agendas and economic conditions. However, their analysis of changes in the English, Scottish, Australian and Danish school systems led them to conclude that NPM was characterized in all four by the following five trends.

- (1) A movement towards school self-management.
- (2) The imposition of greater demands for financial accountability.
- (3) An increase in the extent of consumer control over what goes on in schools.
- (4) The expectation that schools expose their performance to public scrutiny.
- (5) Increasing pressure for outcome-based assessment of pupils, heads and teachers.

Gleeson and Husbands (2001, 1) have characterized the management revolution in schools as follows:

The drive to improve indicators of educational performance, and to ensure that teachers are equipped and able to operate in rapidly changing professional environments, is leading to attempts to reorganise, reskill and reculture teaching . . . A common thread in many, especially western, reforms is the increasing focus on the performance of schools, on the capacity of headteachers to measure that performance and on holding teachers more directly accountable for it through both rewards and sanctions.

Into this environment was born England's Performance Management in Schools initiative (see, for example, DfEE, 2000a, 2000b, 2000c and 2000d, and Brown, 2005) in September 2000, which requires schools' governing bodies to ensure that the performance of all teachers (including headteachers) in state-maintained schools is reviewed annually. The initiative has been described as the world's biggest performance management system (see, for example, NAHT, 1999, 21, and Mahony and Hextall, 2001, 182), which might well be true, considering that it covers approximately 18,000 primary schools, 3,500 secondary schools, 1,100 special schools, 500 nursery schools, 23,000 headteachers, 400,000 teachers and an unknown number of ancillary staff.

The principal aims of this chapter are concerned with the measurement aspects of performance management, and include a consideration of some of the:

- purposes of performance measurement in the context of England's primary school teachers and managers;
- theories and assumptions underpinning these various purposes;
- problems that can arise when these theories and assumptions are applied to the particular context of primary education in England; and
- tensions that are associated with a lack of compatibility between some of the various purposes.

A triangulation of research methods was used, so as to enable a variety of both theoretical and practical perspectives to be considered. Methods included:

- in-depth audiotaped interviews, with thirty headteachers, eight deputy headteachers, thirty-two class teachers and ten school governors, from a diverse range of England's primary schools;
- documentary analysis of official government publications, schools' internal performance-related documents and a variety of both theoretical and practitioner-oriented publications; and

- participant observation (in addition to his role as a researcher, the author of this chapter is also a primary school governor, and one of the responsibilities of this particular position involves conducting the performance management and measurement arrangements of one of England's primary school headteachers).

Purposes of measuring the performance of primary school teachers

The fact that all teachers and heads are required to have a set of between three and six annual objectives (see, for example, DfEE, 2000a, 6, and 2000d, 7) as part of their performance management arrangements indicates that one of the purposes of measuring their performance is to focus their attention onto specific aspects of their work. This particular purpose is encapsulated by the frequently quoted aphorism “What gets measured gets done/attention” (see, for example, Eccles, 1998, 27, Armstrong, 2000, 52, and Latham, 2004, 109).

According to the British government's Department for Education and Skills (DfES, previously known as the Department for Education and Employment, or DfEE), which introduced the Performance Management in Schools initiative, at least one of the objectives of teachers should be concerned with “pupil progress” (DfEE, 2000a, 14), and this – i.e. improving pupils' academic progress – can be considered to be a second purpose of measuring the performance of teachers. As one of their other objectives is being “expected to cover professional development” (DfEE, 2000a, 14), a third purpose of measurement is to improve teachers' professional development arrangements by, first, identifying or evaluating individuals' professional development needs and then assessing the extent to which improvements in those particular aspects of their performance subsequently occurred.

The requirement that an individual teacher's objectives should “relate to the objectives in the school development plan” (DfEE, 2000a, 6) indicates that a fourth purpose of measuring teachers' performance is to facilitate the implementation of organizational (or, in this particular context, whole-school) strategy. A fifth and final purpose that will be considered in this chapter is the highly controversial one of enabling decisions to be made about teachers' performance-related pay increases.

The extent to which these five purposes are achieved or realized in practice tends to be highly variable, for at least two major reasons. First, some of the

theoretical assumptions that underpin the various purposes of performance measurement seem to run into difficulties when applied to the particular context of primary education in England. Second, there appear to be a variety of tensions, conflicts and compatibility problems between some of the various purposes. The first of these matters will now be addressed for each of the five purposes that were outlined above, before moving on to a consideration of the compatibility problems that are associated with some of them.

Measuring for the purposes of focusing attention: theory and practice

Although the concept of “management by objectives” was initially developed by Peter Drucker (see, for example, Drucker, 1954), the idea of incorporating a small number of specific objectives, targets or goals into teachers’ performance management arrangements appears to be based on goal-setting theory, which is usually associated with the work of Edwin Locke and Gary Latham (see, for example, Locke, 1968, Locke and Latham, 1990, and Latham, 2004). The essence of the theory centres on the assumption that setting individuals difficult or challenging specific goals leads to significantly higher performance than easy goals, no goals or vague goals, such as urging people to do their best (Latham, 2004, 107). Although attempts to apply goal-setting theory to the context of primary education have resulted in a certain amount of improvement in the focusing of teachers’ attention onto aspects of their work that are considered to be of particular importance, at least three kinds of problem have occurred.

Problems associated with the meaning of “performance” and “objectives”

A number of writers, including, for example, Williams (2002, 70), have drawn attention to the confusion and disagreement that exists concerning whether the meaning of “performance” is more to do with results/outputs/outcomes or behaviours/processes. Consequently, the literature is divided on whether the formulation and subsequent assessment of performance management objectives should be concerned primarily with the measurement of results/outputs/outcomes, behaviours/processes, either or both. For example, Bacal (1999, 61) defines an objective as “a statement of a specific result or outcome the employee is expected to create or contribute to”. Torrington *et al.* (2003, 93) go even further in arguing that “the focus

must be on what is achieved: results – not behaviours – are what count”. Armstrong (2000, 33–4), however, claims that “objectives” and “goals” are interchangeable terms and can be either “targets” or “tasks” (or “projects”), with targets being quantifiable results to be attained and tasks or projects being units of work to be completed by specified dates in order to achieve the desired results.

With regard to the context of England’s primary schools, the issue of whether objectives should be tasks or targets seems to be particularly problematic, in that the views of the DfES (a ministerial government department) and OFSTED (the Office for Standards in Education – a non-ministerial government department that is responsible for inspecting the standards of schools and teaching in England) appear to be inconsistent. OFSTED, for example, was critical of the fact that

there was scope for improvement in at least half of them. There was a tendency, for example, for objectives to be activities or tasks rather than targets. (OFSTED, 2002, 3)

Nonetheless, although some of the example objectives appearing in documents published by the DfES (DfEE, 2000a, 15–18; DfEE, 2000d, 19–20) are also based around targets/results/outcomes, others are more task-/behaviour-/process-oriented, and yet others appear to be a combination of both. Such confusion has often resulted in variable amounts of tension and conflict between appraisers and their appraisees, with the latter typically arguing for objectives that are more oriented towards behaviours/processes and the former preferring results-/outcome-type objectives, which tend to be more easily measurable. OFSTED (2002) offers the following reason as to why both teachers and heads, when performing the role of appraisee, tended to resist objectives that were based around targets/results/outcomes:

Quite clearly, in some schools, the teachers were reluctant, as were the headteachers, to set quantified objectives for fear of failing to achieve them . . . It was evident that, in some of these cases, the headteachers were reluctant to commit themselves to anything quantifiable for fear of failing to meet the objective, with its predicted impact on pay. (OFSTED, 2002, 9, 11)

Additional problems relating to the definition and meaning of performance management objectives include whether an objective should represent only one or a series of behaviours or results. Bacal (1999, 61), for example, argues that objectives should be “as specific as possible”, “short, to the point, and direct” and that each objective should be focused on a single job responsibility.

Moreover, although the DfES has also argued that “objectives need to be clear and concise” (DfEE, 2000d, 7), one of its example objectives that appears in the same document seems to be anything but:

To raise achievement of underachieving pupils by 2–3% during the year by examining all available data on pupil performance with respect to ethnicity, English as a second language and home circumstances and to use the information to develop strategies to remedy this under performance by December 2000 including a robust monitoring system against which future progress can be measured. (DfEE, 2000d, 20)

One of the consequences of including a subset of additional goals, targets or tasks under the umbrella of a single objective is that, whilst some teachers and heads tend to be given a set of objectives comprising as few as three clearly identifiable specific outcomes or activities, others end up with as many as fifteen or twenty, which seems to defeat the whole measurement purpose of focusing individuals’ attention.

Problems associated with setting objectives that are difficult or “challenging”

In line with Locke and Latham’s theory that “difficult specific goals lead to significantly higher performance than easy goals, no goals or vague goals such as urging people to do their best” (Latham, 2004, 107), the DfES has asserted that teachers’ performance management objectives “should be challenging – neither too comfortable nor dauntingly unrealistic” (DfEE, 2000a, 6). However, although this line of reasoning might well be valid for contexts in which the price of failure is relatively low, it is arguably less so in situations where individuals are fearful of the possible negative consequences of failure (see, for example, Fitz-Gibbon, 1996, 74). Human motivation is not homogeneous, and, while some people in certain workplace environments tend to rise to the challenge of difficult goals, others seem to be crushed at the outset by fears regarding the possible consequences of failure. Furthermore, the particular form of performance management that has been adopted in England’s primary schools seems especially capable of injecting a sense of fear into the minds of both heads and teachers, for at least three reasons.

First, because much of the data on pupils’ academic performance are now published in newspapers and on the internet,¹ many teachers and, in particular, heads tend to fear such things as public humiliation should they

¹ See, for example, www.dfes.gov.uk/performanceables.

fail to achieve the required percentages comprising their pupil progress objectives. Second, because the performance management system that has been chosen for schools incorporates an element of performance-related pay, some heads and teachers genuinely fear the consequences of failure on their pay packets, especially as the DfES urges school governors to base decisions on changes to headteachers' pay on, amongst other things, the extent to which they are judged to have achieved their objectives (DfES, 2003a, 39). Third, because of anxieties about the negative publicity surrounding so-called "failing schools", schools going into "special measures" and the possibility of eventual closure, many headteachers tend to be fearful that a series of failures concerning the achievement of their performance management objectives might eventually result not only in their losing their present job but also their entire future careers being put on the line.

Goal-setting theory, environmental uncertainty and control over performance

A third problem associated with using goal-setting theory for the purposes of focusing people's attention, when applied to the context of England's primary school teachers, is the one of environmental uncertainty. Indeed, Locke and Latham themselves caution about there being "exceptions to the principles" (of goal-setting theory), one of which revolves around the unpredictable nature of certain workplace environments:

Among the biggest impediments to goal setting is environmental uncertainty . . . As uncertainty increases, it becomes increasingly difficult to set and commit to a long-term goal. (Latham, 2004, 114)

The plethora of government change initiatives shortly before and since the introduction of performance management into primary schools has contributed towards their environments becoming uncertain, to the extent that a set of objectives that might have seemed appropriate at the beginning of one particular academic year may no longer appear so several months later. In the words of one of the headteachers who was interviewed as part of this research,

Sometimes objectives which seemed important at the beginning of the year are no longer relevant or important by the mid-year review, because other initiatives have been handed down from above and have become an even greater priority than one's initial objectives.

An additional contributor to the problem of environmental uncertainty in primary schools concerns the extent to which teachers actually have control over their pupils' academic progress, owing to a series of factors both inside and outside the school gates. One of the most obvious of these is the disruptive behaviour of some pupils. Others will be considered in the next section.

Measuring to improve pupils' academic progress: theory and practice

The second purpose of measuring teachers' performance that was identified earlier is that of improving their pupils' academic progress. The theory or rationale behind this particular purpose is that setting "pupil progress" objectives for teachers and subsequently measuring the extent of their achievement tends to concentrate their attention on improving children's academic attainment. Although many teachers and their performance managers who were interviewed for this research did indeed believe that pupil progress objectives had led to genuine improvements in at least some of their pupils' academic progress, there were at least three major problems associated with this particular purpose of measuring teachers' performance.

The unintended and dysfunctional consequences of performance measurement

A considerable amount of research evidence acquired in recent years demonstrates that heads and teachers in primary and secondary schools alike have engaged in a variety of unintended and dysfunctional behaviours that tend to give the *appearance* of improved pupil performance but that might not, in reality, represent *genuine* improvement (see, for example, Gray *et al.*, 1999, Visscher *et al.*, 2000, Karsten, Visscher and de Jong, 2001, and Wiggins and Tymms, 2002). Such behaviours include: teaching to the tests or examinations; focusing a disproportionate amount of attention onto only certain pupils (typically those who are judged as being at the borderline of being able to achieve the required grade or level); curriculum narrowing in favour of examined subjects (English, mathematics and science in primary schools); excessive amounts of revision and cramming; grade inflation to give the appearance of improvements in raw examination data; grade deflation during the earlier stages of primary education in order to give a subsequent impression of improvements in pupils' value-added performance; and, in some cases, blatant or outright cheating, where by pupils have on occasion, simply been told the answers to certain questions.

The validity of using measures of pupil attainment to assess teacher performance

Using indicators of pupil attainment, such as examination results, to measure the performance of teachers assumes a simplistic cause and effect linkage but does not take into account the influence of factors other than teachers on pupils' rates of academic progress during any one particular year. Additional influences, however, include school budgets, which tend to affect such matters as the number of pupils per class, whether classes have to be vertically grouped (i.e. classes containing pupils whose age range is greater than one year) and the affordability of, for example, books, computers and teaching assistant time. Perhaps even more problematic is the difficulty of disentangling the effects, on pupils' progress, of their teachers from those of other influences, such as teaching assistants, help from parents at home, parental attitudes to education, the use of private tutors outside school hours, health-related issues and various social and psychological effects, such as family breakdown. Goldstein (2001) has also drawn attention to the limitations of using pupil performance data for the purposes of judging schools and teachers.

The availability of appropriate information

A third problem associated with attempts to link the academic performance of pupils to the performance of their teachers concerns the availability of appropriate information, something that the originators of goal-setting theory recognized as a major impediment to the effectiveness of goal setting:

The information required to set learning or outcome goals may be unavailable. (Latham, 2004, 114)

Indeed, in order to set appropriate academically based annual pupil progress objectives or goals for teachers, value-added information on pupils' academic progress must be available on a year-on-year basis. In reality, however, value-added information for pupils in many schools is available only over a period of three or four years. In fact, some of the documents recommended by the DfES for the very purposes of setting annual pupil progress objectives (PANDA (Progress AND Assessment) reports and the Autumn Package; see, for example, DfEE, 2000d, 7) do not contain the kind of year-on-year value-added information (see, for example, DfES, 2002a and 2002b, and OFSTED, 2004) that would be required to ensure that such objectives were fair and reasonably attainable. Although all primary schools are now able to gain access to year-on-year value-added information (by, for instance, opting into annual testing schemes such as PIPS (Performance

Indicators for Primary Schools); see, for example, Tymms, 1999), many are not aware of how to do so, or indeed choose not to do so.

Measuring for professional development purposes: theory and practice

A third purpose of measuring teachers' performance that was outlined earlier concerns improvements to their professional development. Theoretical justification for the professional development aspect of teachers' performance management arrangements can be found in McGregor's (1960) "theory Y" assumption of motivation in organizations, which argues, amongst other things, that people are intrinsically motivated and value such things as responsibility and development (Nelson and Quick, 1996, 74). Other theoretical assumptions upon which this particular purpose of performance measurement is based are as follows. First, effective identification and communication of people's professional development needs will occur through a combination of self-evaluation, observation by others, and discussions between appraisers and appraisees. Second, this will lead to individuals taking appropriate professional development courses and training, which will in turn result in improvements to their professional practice. Although many of the teachers who were interviewed as part of this research *did* believe that their performance management and measurement arrangements had resulted in improvements to their professional development, others did not, for reasons that are considered below.

Identification and communication of professional development needs

Because teachers sometimes feared that measurement or assessment of their teaching performance through observations was being used for purposes other than, or in addition to, professional development (such as for decisions relating to changes in their pay), they occasionally decided to conceal their development needs or relative weaknesses. Also, and mainly owing to unavoidable problems relating to subjectivity and reliability, teachers and their team leaders sometimes disagreed about what they considered to be the teacher's most pressing professional development needs, which often led to tension and conflict. As one teacher put it,

The head decided that *this* was what the teacher needed in the way of experience and training – but it wasn't necessarily what the teacher thought was appropriate or what the teacher felt was needed.

Availability of good-quality professional development training

Implicit in beliefs that performance measurement leads to improvements in teachers' professional development are assumptions that good-quality professional development courses or training are available. A large minority of those teachers who participated in this research, however, were not impressed with the external courses that they had taken:

The quality of courses – some of them, you think: “Well this was money for old rope.” You come away and think: “Well, did I learn anything from that?”

Although teachers were generally happier with the quality of their in-house professional development arrangements, overall levels of satisfaction tended to be variable:

A lot of professional development targets can be done in-house but it depends on the mix of the staff – their skills and experience.

Financing professional development

Many teachers were concerned that measurements or evaluations of their performance might not lead to improvements in their professional development arrangements because of the shortage of money that was available for professional development courses and supply cover:

There is more and more an issue of courses being available but no money to attend them . . . You're paying out £176 or whatever it is now for a supply teacher and then you've got to pay out another 50 or £90 for a course fee – well, we can't do it.

Indeed, some teachers worried that the problem of inadequate funding for professional development was discrediting the Performance Management in Schools initiative:

If you're linking performance management to teachers' professional development, then where's the money for training? There *is* no money for training. *Quality* professional development takes time and it takes money. If you're setting up something which purports to enable people to develop their careers and then it doesn't deliver, the system falls into disrepute.

Measuring for strategy implementation purposes: theory and practice

Performance measurement and management frameworks that attempt to align the interests of employees at all levels with those of their employing

organization are usually based on a cascading model, and have been described by the likes of Williams (2002, 78–82) and Kaplan and Norton (2001, 72–3). Essentially, an organization's stated mission or vision is typically translated into a set of goals and measures, which are then cascaded down into divisional or departmental goals. These are, in turn, further cascaded down the management hierarchy into a set of individual targets or personal objectives, the extent of achievement of which is measured or assessed on a regular basis, in order to achieve – in the words of Kaplan and Norton (1996, 199) – “strategic alignment from top to bottom”.

In practice, however, achieving consistency of goals between different management levels and between the various stakeholder groups within an organization is often difficult, not least because, as Williams (2002, 79–80) points out, different stakeholders have different expectations of performance, and members of different departments, work teams and so on may set goals consistent with their interests as they see them. Williams thus considers the achievement of goal consistency throughout an organization to be one of the main challenges for performance management systems. Indeed, some theoretical perspectives, including pluralist frames of reference (see, for example, Fox, 1973), argue that organizations consist of different interest groups, each with their own legitimate interests, and, because of this, tension and conflict between them becomes almost inevitable and unavoidable.

Another theoretical consideration relating to this particular purpose of performance measurement is agency theory (see, for example, Brickley, Smith and Zimmerman, 1997, 154–8). An agency relationship consists of an agreement under which one party, the principal, engages another party, the agent, to perform some service on the principal's behalf. However, problems often occur, because agents, who frequently have their own agendas, sometimes behave in ways that increase their own utility at the expense of their principals. Because the incentives of principals and agents are frequently not aligned naturally, a variety of mechanisms have been designed and developed to increase the likelihood that agents will not act in their own self-interest when this would be to the detriment of their principal's interests. Various performance management and measurement frameworks are examples of such mechanisms.

Arguably, in relatively small organizations such as primary schools, it should be easier to achieve goal alignment between teachers and their head or the whole school than in much larger organizations, because the number of management levels involved is relatively small, as is, therefore, the number

of principal–agent relationships involved in the process. However, amongst those teachers and heads who took part in this research, the extent to which performance measurement achieved the purpose of strategy implementation through goal alignment tended to vary, sometimes considerably, between schools because of frequent difficulties in gaining agreement between appraisers or team leaders and their appraisees concerning the content of the former's performance management objectives. As one deputy headteacher put it,

Teachers need to make sure *they're* getting something out of it . . . , getting training, getting support or whatever . . . It can't just be the *school* getting something out of it. It's gotta be a two-way thing.

Whilst some heads seemed able to balance skilfully and successfully the needs and interests of their teachers with those of their school as a whole, others tended to experience variable degrees of tension and conflict in their attempts to do so. This matter will be considered in more detail in a later section.

Measuring to implement performance-related pay: theory and practice

The fifth and final purpose of measuring the performance of teachers that will be considered in this chapter is the one that tends to arouse most controversy, namely that of establishing a link between their performance and changes to their pay. A theory that could be used to justify the adoption of performance-related pay schemes is McGregor's (1960) "theory X", which, in contrast to his "theory Y" (considered earlier), assumes that people tend to be indolent, lacking in ambition, disliking of responsibility, self-centred, indifferent to organizational needs, resistant to change and gullible. McGregor believed that people should be treated differently depending on whether they appeared to engage in patterns of behaviour that were predominantly "theory X" or "theory Y". Thus, it could be argued, for example, that performance-related pay might motivate "theory X" individuals to improve their levels of performance, whereas such extrinsic incentives would probably be less necessary for the more intrinsically or self-motivated personalities of the "theory Y" type.

Expectancy theory (see, for example, Vroom, 1964) might also be used to justify performance-related pay. It assumes that the tendency of a particular

incentive to motivate a certain individual to work harder is dependent upon the extent to which:

- the incentive is valued by the person;
- the person believes that the extra effort involved will lead to the levels of performance improvement that are required to achieve the incentive; and
- the person believes that the improved performance subsequently achieved will actually result in the incentive being awarded by the organization's management.

Thus, if a particular employee sufficiently values the amount of performance-related pay that is potentially available, and believes that an acceptable amount of extra effort will result in the required levels of performance being achieved, and is confident that, as a result of the improvements in performance, the performance-related payment will actually be made, then he will be motivated to make the extra effort that is required.

A problem with both “theory X” and expectancy theory, however, when applied to the context of primary education, is that most of the teachers and heads who were interviewed for this research did not consider the amount of money available for performance-related purposes to be sufficiently large to act as a particularly tempting incentive. Furthermore, virtually none of them felt that performance-related pay influenced their behaviour at work, primarily because they tended to be more intrinsically than extrinsically motivated by their jobs. These particular findings broadly support those of Wragg *et al.* (2004, 86).

Measurement problems associated with subjectivity and reliability

A major problem associated with effectively implementing a system or scheme of performance-related pay for primary school teachers and heads concerns the sheer difficulty involved in actually *measuring* their performance. In addition, for instance, to the measurement difficulties that were considered earlier, such as the confusion surrounding the definition or meaning of “performance”, and the validity of linking measures of pupils' academic progress to measures of their teachers' performance, there are additional problems concerning the reliability and subjectivity associated with several aspects of teachers' performance measurement arrangements.

For example, many of the teachers who participated in this research tended to be concerned about at least one of the following subjectivity-/reliability-related measurement problems: the quality of relationships influencing judgements of performance; the appropriateness and professional competence of

some appraisers, in that teachers' performance was occasionally assessed by someone who was not considered to have a sufficient amount of expertise in the particular aspect of their performance that was being measured; an inconsistency of standards adopted by the various team leaders who were appointed to evaluate teachers' performance; and the frequency with which performance was measured, in that annual evaluations of teachers' teaching performance, for example, were sometimes based upon as little as a single lesson observation, which tended to represent less than 0.5 per cent of their annual teaching time.

The use of multiple appraisers

The use of multiple appraisers might help to mitigate the above measurement difficulties that are associated with subjectivity and reliability problems. Schemes of multiple appraisal can take a variety of forms, including, for instance, combinations of two or more of the following: assessment by superiors, subordinates, colleagues at the same level, customers and self-appraisal. A variety of methods, techniques and systems have been proposed in recent years that enable a more balanced perspective of performance to be acquired. One such idea is the performance prism framework (see, for example, Neely, Adams and Kennerly, 2002), which, although developed for the purpose of measuring organizational performance, could probably be adapted for the measurement of individual performance. The performance prism pays particular attention to the inclusion of a diversity of stakeholders in the measurement of performance, including customers, employees, investors, suppliers and the community. Adapting the performance prism framework to help measure the performance of headteachers might be particularly worth considering, especially as performance evaluations of heads should, arguably, be closely linked with measurements of the performance of their organization (i.e. school) as a whole. Another concept associated with multiple appraisers to emerge in recent years is that of 360-degree feedback (see, for example, Lepsinger and Lucia, 1997, and Tornow, 1998), which also involves the assessment of individuals' performance by a variety of other people, including their superiors, peers, subordinates and customers.

Although a certain amount of self-appraisal has long been an important aspect of individual performance measurement, it tends to suffer from the same kinds of accuracy and subjectivity problems that occur when employees are appraised only by their line manager. Williams (2002, 165), for instance, draws attention to work conducted by himself (Williams, 1989) and Fletcher

and Baldry (1999) that showed that employees tend to rate their own performance more highly than do their co-workers and managers. Teachers taking part in this research, however, often tended to be more critical than their appraisers of their own performance.

Although the notion of using the perspectives of customers to help evaluate the performance of heads and teachers might, at first, seem novel, it is not. OFSTED (see, for example, OFSTED, 2003a, 31, and 2003b, 30–1), for instance, solicits the views of pupils and parents as part of its framework for judging the performance of primary schools. And educational researchers, such as Rudduck and Fielding (see, for example, Rudduck, Chaplain and Wallace, 1996, and Fielding, Fuller and Loose, 1999) have written extensively about the value of “pupil voice” as a means of helping to improve the performance of schools.

Particular problems associated with managers being assessed by their subordinates, according to Bernardin (1986; summarized in Williams, 2002, 166), are that subordinates often lack the information or skills required to make valid ratings, may be inexperienced as appraisers, may not have been trained to make accurate ratings, may sometimes inflate ratings in order to avoid retaliation from superiors, may harshly rate managers who are demanding and may favour popular rather than effective managers. Bernardin also argues that managers might be tempted to focus too much attention on pleasing their subordinates at the expense of wider organizational needs, and that some managers may avoid organizations that use subordinate ratings, thus leaving such organizations with recruitment and retention difficulties. The inclusion of assessment by subordinates could thus, if handled inappropriately, have a negative impact on both the effectiveness of managers and the performance of their organization as a whole.

In summary, then, although the idea of using a variety of stakeholder groups to measure the performance of primary school heads and teachers would provide a more balanced perspective of their performance, these benefits need to be weighed against the obvious costs of time and money.

Measuring overall performance

In order to be able to make decisions about performance-related pay, those who appraise heads and teachers are expected to measure or evaluate their *overall* performance. In the words of the DfES,

The purpose of the [headteacher’s performance review] meeting is to . . . reach a judgement about the overall performance of the headteacher during the year,

including achievement against previously agreed objectives . . . The governors will form a view at the end of the process whether to recommend to the appropriate committee a pay award on the leadership scale. (DfES, 2003a, 39)

The notion of *overall* performance, however, in the context of primary school teachers and heads, is extremely problematic. For example, in over 100 pages of documentary advice and guidance to governors and headteachers concerning the performance management arrangements of heads (DfEE, 2000d; DfES, 2003a, 2003b), the DfES mentions virtually nothing about precisely *how* a head's overall performance should be measured, or, indeed, what criteria comprise "overall performance". Vague reference is made, in one of the documents (DfES, 2003b, 24), to the usefulness of the Headteachers' Standards Framework in evaluating headteachers' overall performance, and the video (DfES, 2003c) that accompanies this document, in a section about assessing headteachers' overall performance, states, "The School Teachers' Pay and Conditions Document, which is produced annually, sets out the core duties of the headteacher." This document,² however, lists approximately forty professional duties required of a headteacher,³ several of which are broken down into a number of subdivisions. The feasibility of two or three school governors fairly and accurately evaluating, on an annual basis, the extent to which their headteacher accomplishes these many duties seems remote at best.

The *Teachers' Standards Framework* document (DfES, 2001) contains forty-seven standards and role expectations of headteachers, which have since been revised and now appear in the *National Standards for Headteachers* publication (DfES, 2004). This more recently published document describes the core purpose of a headteacher and outlines the knowledge, professional qualities and actions that are expected of heads. The total number of these is 149, and the *meaning*, let alone measurability, of many appears to be very unclear.

Probably as a result of the difficulties associated with defining and measuring headteachers' overall performance, a large proportion of the school governors who were interviewed for this research made no particularly rigorous attempts to do so. Consequently, decisions concerning the "performance-related" pay increases of primary school headteachers are often based around criteria that are more related to affordability than they

² The 2004 edition of this document can be viewed online at www.teachernet.gov.uk/_doc/7102/ACF3966.doc.

³ See pages 77–82 of the 2004 edition.

are to performance. Even when performance considerations *do* influence governors' decisions about changes to their headteachers' pay, frequently they merely take the form of judgements of the extent to which heads have achieved their (typically) three or four annual performance management objectives, which are often only tenuously related to notions of their *overall* performance. For teachers on the lower pay scale, performance-related pay, in practice, does not seem to exist, in that, although they are allowed, in theory, to progress zero, one or two points up the scale during the course of any particular year, all those who were interviewed were awarded one and only one point in each of the years since "performance-related" pay had been introduced.

Conflicting purposes of performance management and measurement

As Torrington *et al.* (2003, 89) point out, performance management

is not a neat management bag of tricks containing tried and tested methods that can be slotted in to any situation. It is certainly not a set of standard techniques, like double-entry book-keeping or critical path analysis. Like most of the more effective management practices, it is an APPROACH or cultural style that can be adopted in any organisational setting.

Moreover, because, as demonstrated throughout this chapter, performance management and measurement systems tend to draw upon and incorporate a variety of theories, assumptions and purposes that do not always appear to sit easily with one another, tensions, conflicts and compatibility problems between them are perhaps inevitable and unavoidable. Some of these will now be considered in the context of England's primary education system (though, as Williams, 2002, 219–24, points out, tension and conflict between the various purposes of performance management and measurement are by no means confined to the teaching profession).

The first kind of tension to be considered is that which arises when performance measurement is used to identify teachers' professional development needs but also as a means of determining whether they get a performance-related pay rise. Before performance-related pay was introduced into the teaching profession, several of the teachers who participated in this research claimed that they were willing to be more open and honest with their head or other team leader about what they themselves considered to be their professional development needs. As one teacher put it,

There was no pressure put on anybody and there was a lot of honesty – and so people would say, “Actually, I’m really worried about teaching so and so,” which you never have them say in performance management.

With the introduction of performance-related pay, however, some teachers became worried that confessions of relative professional weakness might subsequently be used by heads as a reason to avoid awarding them a performance-related pay rise. In the words of one of the deputy headteacher interviewees,

Performance management can discourage openness and honesty, because, as soon as you admit weakness, you’re giving ammunition to someone who doesn’t want to give you a recommendation for a pay award.

Another teacher provided a specific example of how she had deliberately attempted to conceal her professional development needs when her teaching performance was being assessed during a lesson observation by her head:

Like all observations, you go all out for whatever is being observed, but it’s not as if I’m going to be using all those every day in normal life. But as the objective was gonna be measured on just one observation, well, I made sure I delivered what she wanted on that day.

A second kind of tension is one that occurs when the purposes of performance measurement include both performance-related pay and a desire to improve pupils’ levels of academic attainment, by setting their teachers difficult or challenging pupil progress objectives that, according to Locke and Latham’s goal-setting theory (outlined earlier), “lead to significantly higher performance than easy goals, no goals or vague goals”. In practice, however, teachers and their team leaders often have very different ideas about whether the content of a particular objective that has been proposed is considered to be sufficiently “challenging”, not least because

governing bodies or those whom they delegate to deal with pay, should decide whether to award performance pay in the light of a review of performance against previously agreed objectives. (DfEE, 2000d, 8)

Compatibility problems between these two particular purposes of performance measurement have resulted in noticeable amounts of tension and conflict between both teachers and their team leaders and heads and their performance management governors.

A third type of tension is one between the purposes of focusing teachers’ attention and, again, performance-related pay. This occurs because, as

outlined earlier, decisions on changes to teachers' and, especially, heads' pay are supposed to be influenced by evaluations of their *overall* performance, which necessarily requires the assessment of numerous criteria. Each of these criteria could be considered as additional objectives that have to be satisfied in order for the individual to receive his/her performance-related pay rise. However, pursuit of these many additional objectives would almost certainly divert attention away from the three or four formally required performance management objectives the very purpose of which is to focus attention onto those aspects of their work that are deemed to be of greatest priority. In other words, measuring *overall* performance for the purposes of determining performance-related pay rises appears to undermine the other purpose of focusing heads' and teachers' attention onto *specific* areas of their work.

Not all tensions result from decisions to include performance-related pay as part of a performance management or measurement system. A fourth potential area of conflict, for instance, is that between the purposes of facilitating the implementation of organizational or whole-school strategy, on the one hand, and improving the professional development of teachers, on the other. For example, the DfES has advised that teachers' performance management objectives should relate to the strategic objectives of the whole school:

The team leader's role is to ensure that . . . the individual teacher's objectives relate to the objectives in the school development plan. (DfEE, 2000a, 6)

Nevertheless, the same document also asserts that the career aspirations of the teacher will help in setting his/her professional development objectives (DfEE, 2000a, 14). The potential for confusion and conflict implied by such ambiguity was recognized by a number of the teachers who participated in this research:

I'm concerned that the balance is right in that . . . yes, the headteacher's thinking of the good of the school and what he wants to see happening at the school, but he's also thinking about that teacher's development, and that he should be supporting that teacher's development. I think it's a very very difficult job to get that balance right and I'm concerned that there might be some schools where it's not happening.

The following teacher appeared to be working in one of those schools:

My objectives were more about what I could do to contribute to the school rather than what I could do to improve my own professional development . . . I thought there's a whole load of work to do here – which I don't mind. But it's all work that's going to contribute to the life of the school rather than make *me* a better teacher.

Summary, conclusions and policy implications

This chapter has considered some of the purposes of performance measurement along with associated theories, assumptions, problems and tensions in the context of England's primary school teachers and managers. Two principal conclusions can be drawn. First, the extent to which some of the various purposes of performance management and measurement can be achieved or realized in practice appears to be highly context-dependent. Second, some of the purposes of performance management and measurement systems do not appear to be particularly compatible with one another. Because of this, managers who are intending to implement a new system should be clear at the outset about their reasons for doing so, should be aware of the potential for tension and conflict between some of its purposes, and should strongly consider imposing limitations on the number of its stated purposes. As Williams (2002, 219–21) observes, a particular performance management or measurement system might have too many purposes, with the consequent risk that none of them comes to be well served.

References

- Armstrong, M. (2000). *Performance Management: Key Strategies and Practical Guidelines*, 2nd edn. London: Kogan Page.
- Bacal, R. (1999). *Performance Management*. New York: McGraw-Hill.
- Bernardin, H. (1986). Subordinate appraisal: a valuable source of information about managers. *Human Resource Management*, 25, 421–39.
- Brickley, J., Smith, C., and Zimmerman, J. (1997). *Managerial Economics and Organizational Architecture*. Boston: Irwin/McGraw-Hill.
- Brown, A. (2005). Implementing performance management in England's primary schools. *International Journal of Productivity and Performance Management*, 54(5–6), 468–81.
- DfEE (2000a). *Performance Management in Schools: Performance Management Framework*. London: Department for Education and Employment.
- (2000b). *Performance Management in Schools: Guidance Note*. London: Department for Education and Employment.
- (2000c). *Performance Management in Schools: Model Performance Management Policy*. London: Department for Education and Employment.
- (2000d). *Performance Management: Guidance for Governors*. London: Department for Education and Employment.
- DfES (2001). *Teachers' Standards Framework*, DfES/0647/2001. London: Department for Education and Skills.

- (2002a). *The Autumn Package 2002 Pupil Performance Information Key Stage 1*. London: Department for Education and Skills.
- (2002b). *The Autumn Package 2002 Pupil Performance Information Key Stage 2*. London: Department for Education and Skills.
- (2003a). *Performance Management 2003: Support Guide for Governors and Headteachers*, DfES/0568/2003. London: Department for Education and Skills.
- (2003b). *Performance Management 2003: Support for Governors and Headteachers: Workbook*, DfES/0535/2003. London: Department for Education and Skills.
- (2003c). *Performance Management 2003: Support for Governors and Headteachers: Video 2 – Headteacher Performance Review Meeting*, DfES/0534/2003. London: Department for Education and Skills.
- (2004). *National Standards for Headteachers*, DfES/0083/2004. Nottingham: Department for Education and Skills.
- Drucker, P. (1954). *The Practice of Management*. New York: Harper and Row.
- Dunleavy, P., and Hood, C. (1994). From old public administration to new public management. *Public Money and Management*, 14(3), 9–16.
- Eccles, R. (1998). The performance measurement manifesto. In *Harvard Business Review on Measuring Corporate Performance*, 25–45. Boston: Harvard Business School Press.
- Fielding, M., Fuller, A., and Loose, T. (1999). The central place of pupil voice in primary school improvement. In Southworth, G., and Lincoln, P. (eds.), *Supporting Improving Primary Schools: The Role of Heads and LEAs in Raising Standards*. London: Falmer.
- Fitz-Gibbon, C. T. (1996). *Monitoring Education: Indicators, Quality and Effectiveness*. London: Continuum.
- Fletcher, C., and Baldry, C. (1999). Multi-source feedback systems: a research perspective. In Cooper, C., and Robertson, I. (eds.), *International Review of Industrial and Organizational Psychology*, 149–93. Chichester: John Wiley.
- Fox, A. (1973). Industrial relations: a social critique of pluralist ideology. In Child, J. (ed.), *Man and Organization*, 185–233. London: Allen and Unwin.
- Gleeson, D., and Husbands, C. (2001). *The Performing School: Managing Teaching and Learning in a Performance Culture*. London: Routledge Falmer.
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: scope and limitations. *British Educational Research Journal*, 27(4), 433–42.
- Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S., and Jesson, D. (1999). *Improving Schools: Performance and Potential*. Buckingham: Open University Press.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Kaplan, R. S., and Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- (2001). *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Boston: Harvard Business School Press.
- Karsten, S., Visscher, A., and de Jong, T. (2001). Another side to the coin: the unintended effects of the publication of school performance data in England and France. *Comparative Education*, 37(2), 231–42.
- Kettl, D. F. (2000). *The Global Public Management Revolution: A Report on the Transformation of Governance*. Washington, DC: Brookings Institution Press.
- Lane, J. E. (2000). *New Public Management*. London: Routledge.

- Latham, G. P. (2004). Motivate employee performance through goal-setting. In Locke, E. A. (ed.), *The Blackwell Handbook of Principles of Organizational Behaviour*, updated edn. Oxford: Blackwell.
- Lepsinger, R., and Lucia, A. (1997). *The Art and Science of 360-degree Feedback*. San Francisco: Jossey Bass.
- Locke, E. A. (1968). Towards a theory of task performance and incentives. *Organizational Behaviour and Human Performance*, 3(2), 157–89.
- Locke, E. A., and Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.
- McGregor, D. M. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.
- Mahony, P., and Hextall, I. (2001). Performing and conforming. In Gleeson, D., and Husbands, C. (eds.), *The Performing School: Managing Teaching and Learning in a Performance Culture*, 174–91. London: Routledge Falmer.
- Moos, L., and Dempster, N. (1998). Some comparative learnings from the study. In MacBeath, J. (ed.), *Effective School Leadership: Responding to Change*, 98–111. London: Paul Chapman.
- NAHT (1999). *Response to the Green Paper "Teachers: Meeting the Challenge of Change" and Technical Consultation Paper on Pay and Performance Management*. Haywards Heath: National Association of Headteachers.
- Neely, A., Adams, C., and Kennerly, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Harlow: Financial Times/Prentice Hall.
- Nelson, D., and Quick, J. (1996). *Organizational Behavior: The Essentials*. St Paul, MN: West.
- OFSTED (2002). *Performance Management of Teachers: Report on the Implementation of the Government's New Performance Management Requirements for Teachers, Based on a Sample of Primary, Secondary and Special Schools*. London: OFSTED.
- (2003a). *Inspecting Schools: Framework for Inspecting Schools*. London: OFSTED.
- (2003b). *Inspecting Schools: Handbook for Inspecting Nursery and Primary Schools*. London: OFSTED.
- (2004). 2004 PANDA report anonymous primary school. Available at www.ofsted.gov.uk/publications/index.cfm?fuseaction=pubs.displayfile&id=3760&type=pdf (20 November 2005).
- Osborne, D., and Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Plume/Penguin.
- Rudduck, J., Chaplain, R., and Wallace, G. (1996). *School Improvement: What can Pupils Tell Us?* London: Fulton.
- Tornow, W. (1998). *Maximizing the Value of 360-degree Feedback*. San Francisco: Jossey Bass.
- Torrington, D., Earnshaw, L., Marchington, L., and Ritchie, E. M. D (2003). *Tackling Under-performance in Teachers*. London: Routledge Falmer.
- Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-added Indicators*. London: David Fulton.
- Visscher, A., Karsten, S., de Jong, T., and Bosker, R. (2000). Evidence on the intended and unintended effects of publishing school performance indicators. *Evaluation and Research in Education*, 14(3–4), 254–67.
- Vroom, V. (1964). *Work and Motivation*. New York: John Wiley.
- Wiggins, A., and Tymms, P. (2002). Dysfunctional effects of league tables: a comparison between English and Scottish primary schools. *Public Money and Management*, 22(1), 43–8.

- Wragg, E. C., Haynes, G. S., Wragg, C. M., and Chamberlin, R. P. (2004). *Performance Pay for Teachers: The Experiences of Heads and Teachers*. London: Routledge Falmer.
- Williams, R. S. (1989). Alternative raters and methods. In Herriot, P. (ed.), *Assessment and Selection in Organizations: Methods and Practice for Recruitment and Appraisal*, 725–36. Chichester: John Wiley.
- (2002). *Managing Employee Performance: Design and Implementation in Organizations*. London: Thomson.

Police performance: sovereignty, discipline and governmentality

Paul Collier

Introduction

This chapter considers performance measurement as a form of knowledge that is intimately connected with power. The setting for the study is the measurement of police performance and how performance measurement constitutes and is constituted by knowledge and power. The work of French philosopher Michel Foucault is used to understand how performance measurement has been a tool by which the sovereignty of government over policing has been established, and how performance measurement has been adopted as a disciplinary measure, such that both can be viewed as governmental technologies involving the use of power to constitute knowledge deemed to be important for political purposes.

The context of policing

Policing in England and Wales (the “Home Office” forces) is carried out through forty-three relatively autonomous police forces. Governance of policing is tripartite, comprising the chief constables, a police authority for each force area and the Home Secretary. Chief constables (or commissioner in the case of the Metropolitan Police), as the professional head of each force, have independence from political interference in operational matters. Police authorities are responsible for setting the force budget, determining local priorities and holding chief constables accountable for their performance. The Home Secretary is responsible for the efficiency and effectiveness of policing nationally. Her Majesty’s Inspector of Constabulary (HMIC)

The author would like to acknowledge the support of the ESRC/EPSRC AIM Research initiative under grant number RES-331-25-0010 for this research. The author is grateful for the comments of participants at the PMA workshop held at London Business School in September 2005.

reports to the Home Secretary on the efficiency and effectiveness of each force.

The office of “constable” itself dates back over 700 years, to the Statute of Westminster in 1285, which provided for two constables to be appointed by each hundred to maintain law and order, as agents of the local sheriff. In 1662 Charles II transferred to justices of the peace the power to appoint constables, and the independence of the constable as a crown appointment probably originated with this change (English and English, 1996).

Police constables are not employed under any contract of employment. They hold a public office, a Crown Warrant. The holder of the office bears full personal responsibility for his or her actions and can never claim that he/she was merely acting upon orders if the act is legally or morally questionable. An officer is a member of a police force, although subject to the control of his/her chief officer (English and English, 1996).

The role of constable underlies the notion of constabulary independence, especially in relation to the authority of the chief constable. That there is no “master–servant” relationship is firmly established in law.¹ Lord Justice Denning defined the duty of a chief constable as

to enforce the law of the land . . . but in all these things, he is not the servant of anyone, save of the law itself.²

The independence of police officers from political interference has, however, been harmed by a number of events over the last thirty or so years. Concern about the violation of the rights of suspects in the 1970s led to a Royal Commission on Criminal Procedure in 1981 and subsequently to the Police and Criminal Evidence Act (PACE) of 1984, now the single most important regulator of police powers and responsibilities (Reiner, 2000).

The Scarman inquiry of 1981, following the Brixton riots, was a turning point in the debate about police organization and strategy and the accountability and constitutional position of the police (Reiner, 2000). The urban riots of the early 1980s, the miners’ strikes during the mid-1980s and the anti-poll tax demonstrations of the early 1990s had the effect of altering the perceptions of the public about the police, while miscarriage of justice scandals led to the Runciman Royal Commission on Criminal Justice in 1991.

Over the last twelve years accountability has focused on the management of policing, increasingly referred to as “value for money” or “best value”. Police

¹ *Fisher v. Oldham Corporation* (1930), 2 KB 364.

² *R v. Metropolitan Police Commissioner ex p. Blackburn* (1968), 1 All ER 763.

reforms followed two White Papers (Home Office, 1993, 2001) emanating from broader public sector concerns about performance management.

The emergence of performance measurement in police forces

The emergence of public sector performance measures has been an important element of the new public management and its emphasis on accountability in terms of results (Hood, 1995). There has been a global movement towards the measurement of performance (OECD, 1997), much of which emanates from the notion that government needed to be “reinvented” into a more entrepreneurial form (Osborne and Gaebler, 1992).

Like other public sector organizations in the United Kingdom, a concern with performance affected the police service after the introduction by the Conservative government headed by Margaret Thatcher of the Financial Management Initiative in 1982. The requirement on the police to set objectives and priorities was introduced in 1983 and the requirement to set performance indicators was announced in the Citizen’s Charter in 1991. Her Majesty’s Inspector of Constabulary had started to develop performance indicators in 1986 (Carter, Klein and Day, 1992), although forces were not required to produce performance indicators for HMIC until 1992.

Collier (2006b) notes that there has been a continual change in what is measured, how it is measured and the targets used since 1992. The life cycle of earlier initiatives has been short and the changes made through each successive initiative have been neither incremental, unidirectional nor unambiguous.

At the time of writing the “performance regime” comprises public service agreements (PSAs), a yearly National Policing Plan, statutory performance indicators (SPIs) and the Police Performance Assessment Framework (PPAF). Public service agreements were an outcome of the government’s biannual spending review and aimed to ensure that resources allocated to departments were linked to results. The annual National Policing Plan sets out the strategic policing priorities, including objectives and performance indicators for police forces. A suite of statutory performance indicators has been introduced that, at the time of writing, contains thirty-six indicators and sub-indicators, twenty-three of which are output (arrests, detections) or outcome (public satisfaction) measures. The Policing Performance Assessment Framework is a model of the “balanced scorecard” type that has six “domains”: citizen focus; reducing crime; investigating crime; promoting

public safety; providing assistance; and resource usage. The PPAF is published annually, using a selection of the SPIs and as a set of “radar” or “spider” diagrams that benchmark the performance of each police force against their “most similar” forces on the basis of a number of exogenous values. Performance data are collected by each force and submitted monthly to the Home Office, where it is stored in iQuanta, a large database system that produces 300,000 graphs each month providing multiple views of data on crimes and detections in each force and local command area.

There have been many studies of the dysfunctional consequences of performance indicators (Likierman, 1993; Rogerson, 1995; Smith, 1995a; Fitzgerald *et al.*, 2002; Carmona and Gronlund, 2003). In policing studies, concern has been raised about monitoring some easy-to-measure indicators that provided a traditional view of police work while some crucial areas of policing were not measured (Carmona and Gronlund, 2003). This is particularly important in relation to crime levels and detection rates, which are the focus of police performance measurement. However, financial statistics reveal that only 27 per cent of the national police budget is devoted to crime fighting (CIPFA, 2003).

Performance in the public sector is inevitably a political tool as much as it is a managerial one. Public management cannot be divorced from politics (Gray and Jenkins, 1995) and performance indicators are important in informing the electorate of the activities undertaken by the public sector (Smith, 1995b). Reforms may have a dual purpose, legitimizing the police service to the electorate while encouraging efficiencies of resource use (Hoque, Arends and Alexander, 2004).

For some years the government has been concerned with the relative performance of police forces in terms of the financial resources and the results obtained. This has been the subject of several reports (Audit Commission, 1998a, 1998b, 1999; Home Office, 1999; Public Services Productivity Panel, 2000).

A study of performance measurement in four police forces

This chapter builds on the results of research into police financial and non-financial performance across England and Wales over several years, but emphasizes data collected through case study research with four particular police forces during 2004 and 2005. The four forces are identified as Large, Medium 1, Medium 2 and Small, based on their size relative to the thirty-nine

English police forces. The study took place at both the headquarters and the basic command unit (BCU) levels. The BCU is the main operational unit, usually coterminous with local authority boundaries, within which most operational policing takes place, supported by a headquarters in each force providing specialist support.

The case studies in all four police forces comprised a series of forty-six interviews, each lasting approximately one hour, with chief officers and senior police officers (the ranks of chief inspector, superintendent and chief superintendent) and with lower ranks and police staff. The research was supplemented by interviews within the Home Office.

Access was also obtained to sixteen meetings across the four forces in relation to performance management, lasting up to four and a half hours. The meetings had between twenty and forty-six participants. In addition, the four forces provided more than twenty comprehensive performance measurement documents. The research was supported by desk-based research of publicly available information, primarily from the Home Office, extending over several years.

Most police officers admit that, even five years ago, they believed they were largely unable to impact crime on account of the responsibility of education, health and social services for its socio-economic causes. However, this attitude has shifted, partly because of the increasing emphasis of government on improving performance and holding chief police officers accountable for that performance, with the rewards and sanctions accompanying that. Perhaps more importantly, it is a result of the professional policing experience in New York, where the Compstat model was developed (Bratton, 1998; Maple, 1999). Compstat held local police commanders publicly accountable for their performance by measuring levels of crime and by maintaining a rigid focus on continuous reductions in crime through targeted policing activity.

This attention to performance is not a new development. Before the 1990s performance was considered largely in terms of process, evidenced by the Royal Commissions and other inquiries described earlier in this chapter. Statutory performance indicators were first introduced in 1992, but it was not until 2002 that the government's Spending Assessments for public services introduced PSAs imposing specific performance targets. In policing, this coincided with the development of the iQuanta system to centralize data collection and reporting, and with the introduction of a Police Standards Unit within the Home Office, charged with improving the overall and relative performance of police forces. Through most of the 1990s HMIC had

been the main body inspecting forces, and “league tables” of SPIs had been published annually by the Audit Commission.

As part of this process, the performance focus shifted from force (i.e. headquarters) level to BCU level, at which most operational policing takes place. The Police Standards Unit monitored BCU performance and compared each BCU to its “most similar” BCU group, while HMIC began inspections of BCUs that were independent of inspections of forces. Throughout, at both force and BCU level, there was an increased emphasis on self-inspection and improvements driven at the force level.

Each force had established a performance management department in its headquarters. A great deal of effort went into analysing performance at force level, with information broken down in terms of BCUs, and in some forces down to police station and shift level. This information was routinely produced for senior police managers but was also accessible through the force intranet. National comparisons were also available from the Home Office’s iQuanta system.

As a managerial tool, performance measurement is used to identify areas for improvement: particular crimes that are on the increase, crimes that are difficult to detect, particular locations where there is a “hot spot” of activity, prolific offenders who are responsible for a disproportionate amount of crime or disorder, etc. Each force produces data at force and BCU level that compare performance to targets, identify trends over time and taking into account seasonal factors, and benchmark performance both within the force and to “most similar” forces nationally. “Best practice” information is routinely shared between forces in order to improve performance.

Each of the four forces studied held a monthly or bi-monthly performance meeting to review the performance data. At these meetings, which involved a large number of police managers, BCU commanders were held accountable for their performance. A quarterly visit by a chief officer to each BCU provided a more in-depth analysis of performance.

Different approaches to performance were adopted both between and within forces. In management meetings at BCU level and in “tasking and co-ordinating” meetings, which allocated resources and determined policing priorities, performance measurement was always evident. While some forces and BCUs were concerned with being seen as underperforming by the Home Office, others were concerned with improving their position relative to other forces, while others were concerned simply with improving the trend of their performance over time. However, the focus on performance has been pushed down to most levels of policing.

One of the tensions in policing has been the difference between levels of crime – which, according to all published statistics, have decreased – and the fear of crime, which, according to the British Crime Survey, has increased. This has been linked to the public's concern with antisocial behaviour, defined by the Crime and Disorder Act, 1998, as behaviour that causes or is likely to cause harassment, alarm or distress to one or more people who are not in the same household as the perpetrator. It ranges from football hooliganism, domestic disputes, youths playing on street corners, and alcohol-fuelled disorder after pubs and clubs close in the evening. Much antisocial behaviour is not crime.

The provisions of the Crime and Disorder Act require police forces to work closely with local authorities in developing local police priorities. The Office of the Deputy Prime Minister (ODPM) exercises its power through regional government offices and Crime and Disorder Reduction Partnerships (CDRPs), which are affiliated with local authorities.

Although antisocial behaviour is much higher on the agenda of police forces, a major weakness is the absence of clear measures, although the National Standard for Incident Recording, introduced in 2005, tries to overcome this. In the absence of measures, crime and detections feature far more prominently in police management.

The government's increasing attention to citizen focus and neighbourhood policing (ODPM, 2004; Home Office, 2004a, 2004b) is likely to rebalance performance management such that antisocial behaviour and the fear of crime take on a much higher profile in the future. Collier (2006b) shows the shift in the type of performance measures collected under "best value" towards process indicators, and the shift that has taken place since 2004 from process indicators in favour of outputs (detections) and outcomes (surveys of public satisfaction).

Performance measurement as knowledge and power: introducing Foucault

Performance measurement constitutes a form of knowledge that is used as power by government to control policing. To interpret and understand performance measurement in the context of policing and the power–knowledge relationship, we need a lens through which we can view the complexity of the tripartite structure, the notion of constabulary independence and the emergence of the current method by which police performance

is measured, monitored and managed. We can find a suitable lens for this purpose by drawing on the writings of the French philosopher Foucault (1926–84).

Foucault's writings have been influential in helping to explain power and knowledge in organizations; his interest was in how we became trapped in our own history. The objective of Foucault (1982, 209) was to create a history of the different modes by which human beings are made subjects and to provide "a historical awareness of our present circumstance".

Foucault (1975) argued that power produces knowledge, and that power and knowledge both directly imply the other: there is no power relationship without the constitution of a field of knowledge, nor knowledge that does not constitute power relations, as knowledge extends to reinforce the effects of power. Foucault did not see the effects of power as negative but as productive. To be useful, Foucault argued that individuals have to be both productive and subjective, and he saw power as a strategy for that enterprise. The mechanics of power ensure that bodies become docile, such that discipline becomes the maximization of efficiency. The result is a "subtle, calculated technology of subjection" (1975, 221).

The historical analysis of the birth of the prison in *Discipline and Punish* (Foucault, 1975) located the right to punish in sovereign power residing in the monarchy. The instrument that allowed sovereignty to achieve its aims was obedience to the laws, and so law and sovereignty were inseparable (Foucault, 1978 [1991]).

Foucault (1975) traced the development of sovereign power through the eighteenth-century reform movement to a techno-politics of punishment that increased the effects of power whilst diminishing both its economic and political cost. Discipline as a technology of power derived from three "simple instruments": hierarchical observation, normalizing judgement and the examination.

Hierarchical observation is concerned with "subjection" and "exploitation", based on the geometry of the military camp, which provides visibility and the "subtle partitioning of individual behaviour" (Foucault, 1975, 173) carried out in workshops and factories by "clerks, supervisors and foremen" (174). In the instrument of normalizing judgement, the "normal" is both individualizing and homogenizing by making it possible to measure gaps between individuals. Using this instrument, the norm becomes the standard against which actual performance is measured and, when necessary, corrected, in which desired performance is reinforced by "gratification-punishment" (180). Examination constitutes the external inspection, using the norm as

the standard, through which individuals are differentiated and judged. It establishes visibility by creating a field of surveillance embedded in a network of writing.

A variant of disciplinary power is “bio power”, a form of self-discipline that Foucault (1976 [1990]) described as emerging from both the confessional and psychoanalysis. Bio power is aimed at the subjugation of human bodies while maintaining their productivity and the control of populations by defining what is normal through discourses of psychiatry, medicine, social work, etc. (Ezzamel, 1994).

In his later work, Foucault (1978 [1991]) shifted his attention to the notion of “governmentality”. Foucault’s interest was the historical links between three “movements”: government, population and political economy. The first movement overturns the constants of sovereignty in consequence of the problem of choices of government; the second movement brings about the emergence of population as datum, as a field of intervention and as an objective of governmental techniques; the third movement is the process that isolates the economy as a specific sector of reality, and the science and technique of intervention by government in that field of reality. Population comes to be the ultimate end of all government as government has as its purpose the welfare of the population, the improvement of its condition and the increase of its wealth, longevity, health, etc.

Foucault’s notion of “government” included, but was more than, the state, having various aspects: the government of oneself; the government of souls and lives by the Church; the government of children through pedagogy; and the government of the state by the prince. The prince was external to his principality, and the objective of the exercise of power was to reinforce, strengthen and protect the principality (Foucault, 1978 [1991]).

In developing an understanding of governmentality, Miller (1990) contrasts “political rationalities”, as “the field of statements, claims and prescriptions that sets out the objects and objectives of government” (317), with “technologies of government”, as the body of techniques that makes the objects and objectives of government amenable to interventions, emphasizing the transformative effects that routine calculations can bring about. Government is a problematizing activity, and it is around difficulties and failures that “programmes of government” have been elaborated, reflected in government reports, committees of inquiry, White Papers, etc., and laying claim to a certain knowledge of the sphere or problem to be addressed. Knowledge is central to calculation and intervention (Miller and Rose, 1990; Rose and Miller, 1992). Clegg *et al.* (2002) describe governmentality as being

premised on the active consent and subjugation of subjects, rather than their oppression, domination or external control, such that reflexive self-control makes external sources of surveillance unnecessary.

Importantly, Foucault did not see a transition from sovereignty to discipline to governmentality. Rather, he saw these as coexisting, “a triangle, sovereignty–discipline–government, which has as its primary target the population and as its essential mechanism the apparatuses of security” (Foucault, 1978 [1991], 102).

Knowledge is contained, as we have seen, in accountings for performance, and these accountings are technologies of government that simultaneously constitute sovereignty and act as a means of discipline. A consideration of accountings should not be limited to a financial meaning, however, as non-financial performance measurement comprises the main form of accounting (“a counting”) in policing.

Accounting (or “a counting”) for performance

All government depends on a particular mode of “representation” in which information is not neutral and “technologies” translate thought into the domain of reality through “techniques of notation, computation and calculation; procedures of examination and assessment; the invention of devices such as surveys and presentational forms such as tables” (Miller and Rose, 1990, 8). The specific calculative practices and rationales of accounting are held to be linked to particular conceptions of government while the construction and elaboration of governmental programmes are processes that call upon the calculative practices of accounting. “Accounting” is defined here as a process of attributing financial values and rationales to a wide range of social practices, thereby according them a specific visibility, calculability and utility (Miller, 1990, 316), although, for the purposes of this chapter, “accounting” is not limited to financial measurement.

For present purposes, the writing of performance measures, although not financial, can be considered to be a process of valuing by ascribing non-financial (i.e. numeric) values to performance such as the number of crimes and the number of detections, which over time and through comparison to targets and benchmarks comprise a process of valuation. This is, indeed, the origin of “statistics”, which worked within and for the benefit of a monarchical administration that functioned as a form of sovereignty. This use of “statistics” has been considered as “political arithmetic” (Hacking, 1975, quoted in Gordon, 1991).

We now consider this accounting for performance in the context of the field study of performance measurement in policing. In doing so, we use Foucault's triangle of sovereignty–discipline–government as the lens through which we can understand transitions in police performance measurement over time.

Sovereignty

The tradition and legality of constabulary independence has been eroded by sovereignty, as reforms of policing have been at first operational but since the early 1990s more managerial, being based on concepts of value for money, best value and performance measurement.

The government of policing can be seen as an assertion of sovereign rights over the control of the population through policing, as well as over policing itself. Sovereign power is exercised by requiring an accounting for performance in relation to levels of crime and rates of detection. Thus, power constitutes knowledge (i.e. performance measures as a form of knowledge), and that knowledge reinforces (sovereign) power (Foucault, 1975).

The research revealed tensions, however, between the sovereignty of the Home Office over the efficiency and effectiveness of the police service nationally, as required by the Police Act; the local sovereignty of police authorities to approve the police budget, set local policing priorities and hold chief constables accountable for their performance; and the constitutional sovereignty of chief constables, who are, at law, free of political interference in operational matters. However, the notion of seizure under sovereignty is evidenced in particular by the Home Office's power, through the Police Standards Unit, to intervene in police forces when performance is deemed unacceptable.

There have been further tensions as the ODPM has attempted to usurp some power from police forces, police authorities and the Home Office, through its role in local government, crime and disorder and CDRPs. To some extent at least, the Treasury also exercises a sovereign role through its power over the total amount of funds allocated to policing, and in particular through the public service agreements, in which it establishes targets for police forces to achieve, in return for the funds provided.

Sovereignty has also been impacted by the increasing autonomy of basic command units. First given increased budgetary authority, BCUs then became subject to performance monitoring by the Police Standards Unit, and subject to inspection by HMIC, separate to monitoring and inspection at

the force level. This has made BCU commanders more accountable, but also more able to argue with force headquarters about resources. Limited funds have been directed from CDRPs to BCUs, bypassing the force headquarters, to support antisocial behaviour strategies.

Discipline

Hierarchical observation takes place in the police service predominantly through the rank structure, which enables a large number of officers to be controlled by successively higher ranks. These officers are organized into BCUs, stations and shifts to assist in the observation and control of their activity. The accountings described in the case studies comprise the measurement and reporting of police performance, particularly in relation to crime and detections. Performance measurement provides visibility over the operation of policing and accountability to each successively higher organizational level, and externally to the government. Hierarchical observation also takes place through the visibility inherent in the publication of performance indicators in the Police Performance Assessment Framework via “radar” diagrams, producing a knowledge about policing from which judgements are made.

The normalizing judgement is based on standardization. With thirty-nine police forces in England, the “normal” is based on a grouping of “most similar” forces. Forces are benchmarked with each other, over time and in comparison to targets in order to exemplify “better” and “worse” performance, either in efficiency or effectiveness. These normalizing judgements are supported by a range of rewards, including promotion and transfer (into favoured postings), and sanctions, such as transfer (into less attractive postings) and a range of disciplinary measures. Normative judgement is reinforced by the monitoring and benchmarking of performance measures through the iQuanta system and the intervention powers of the Home Office’s Police Standards Unit into forces deemed to be underperforming. It is also reinforced by examination.

Examination takes place externally and internally. Externally, it is largely the province of HMIC, the inspection process of which reports on the efficiency and effectiveness of each force through what is called the Baseline Assessment, a qualitative assessment of each police force. HMIC also reports on progress towards implementing relevant government initiatives and what has been identified as “best practice” from other forces. This examination is based largely on the written account of performance produced by each force,

then compared with the norms of other forces. The results of the examination are public documents, and can serve to draw the attention of the Police Authority or Home Office to areas for improvement.

Examination is also internal, in the performance meetings that act as the focus of performance reporting, assessment and improvement for each force. Quarterly visits by chief officers to each BCU provide a form of examination, as they compare and question performance and hold BCU commanders accountable for that performance.

The self-examination that precedes external examination encompasses a detailed written self-assessment by each force, and, together with the whole process of internal examination, forms a confessional, or “bio power” (Foucault, 1976 [1990]).

Governmentality

Problematization in government has been evident in relation to both crime and the fear of crime and the economic performance of police forces, and it is around the difficulties and failures in both these fields that “programmes of government” have been elaborated. Governmental technologies aimed at measuring, managing and improving police performance include public service agreements, the annual National Policing Plans, statutory performance indicators, the Police Performance Assessment Framework, the role of HMIC as the inspecting agency, and the “intervention” powers of the Police Standards Unit. The measurement and management of performance is at the centre of these technologies of government, which can be seen as being aimed at political economy through improved police performance.

The whole focus of governmentality in policing is the measurement and management of performance aimed at political economy through improved police performance, defined as the achievement of better results (e.g. lower crime and increased detections) for the funds invested. Political economy is particularly evidenced in the emphasis on the relative performance of police forces, through comparisons based on the “most similar” forces procedure. It is also evidenced in the attention given to linking performance with funding, originally through introducing a scheme of local financial management (Collier, 2001), then with performance/funding relationships (Public Services Productivity Panel, 2000) and activity-based costing (Collier, 2006a).

Government takes place at several levels: by police forces of themselves; by local police authorities; by the (limited) influence of regional government offices and Crime and Disorder Reduction Partnerships; by the Home Office;

and by the Treasury. The increased attention to (or, at least, the rhetoric of) local issues of citizen focus and neighbourhood policing (ODPM, 2004; Home Office, 2004a, 2004b) provides the population imperative, acknowledging the reflexivity between control of the police by the population and control of the population by the police.

Transitions within the triangle of sovereignty–discipline–government

As Foucault (1978 [1991]) suggested, there is no transition from sovereignty to discipline to governmentality. The police case shows that sovereignty and discipline continue, despite attention being given more recently to government technologies of citizen focus and neighbourhood policing. Sovereignty and discipline are themselves technologies of government.

What can be seen, however, is a transition within the triangle of sovereignty–discipline–government, under which each element gains or loses a different emphasis over time. This transition is shown in table 17.1. The table considers how sovereignty, discipline and government have each shifted over three time periods. The first time period – the “performance measurement” period – covers the period beginning when performance indicators were first introduced, during which measurement did not lead to significant management, and when the focus of attention was local and chief constables dominated police forces.

The second period reflects a transition from performance measurement to management. The sovereignty of chief constables was eroded and a significant centralization of power, chiefly in the Home Office, took place, despite competition for sovereignty over policing. Disciplinary mechanisms increased significantly and the power of intervention was introduced as perhaps the most significant method by which police performance could be influenced (at one time or another, eight of the thirty-nine English police forces have been subject to intervention).

The third period is one of renewed citizen focus, heralded by various government reports and reflected in an expansion of performance to include broader outcomes. The exact nature of this third period cannot yet be discerned, but it is likely to see a renewed emphasis on local issues and local governance.

Table 17.1 shows that Foucault’s triangle of sovereignty–discipline–government reflects shifts in each of the elements of the triangle over the three time periods. The table also illustrates the relationship between

Table 17.1: Transition within the triangle of sovereignty–discipline–punishment over three time periods

	Period 1: Performance measurement	Period 2: Performance management	Period 3: Citizen focus
Timescale	From inception of SPLs in 1992	From 2002 introduction of PSAs	Announced in 2004, but likely to take place in future
Sovereignty	Relatively balanced tripartite relationship Unquestioned control of police forces by chief constables	Erosion of unquestioned power of chief constables Greater power to central government: Treasury (through PSAs), Home Office (through PPAF) and “seizure” power of Police Standards Unit; also exercise of power of ODPM in relation to fear of crime and antisocial behaviour Increased independence of BCUs	ODPM involvement through government offices in crime reduction and antisocial behaviour strategies Announced intention to give greater power to local citizens through restructured police authorities
Discipline	Overall concern with process (various Royal Commissions and inquiries) Performance measures varied Central collection of data by HMIC “League table” reporting by Audit Commission HMIC inspections of forces	Publication of PPAF “radar” diagrams showing performance Police Standards Unit “most similar” force and BCU comparisons HMIC inspections of BCUs Increased emphasis on self-assessment	Performance measures shift from input and process measures to output (crime and detections) and outcome (public satisfaction) measures reflected in PPAF
Governmentality	Local focus under chief constable with limited importance of performance measures	National centralization Intervention powers of Police Standards Unit	Increased attention to local issues, especially antisocial behaviour and the fear of crime Government strategies for citizen focus and neighbourhood policing

Table 17.1: (cont.)

	Period 1: Performance measurement	Period 2: Performance management	Period 3: Citizen focus
Power/knowledge	<p>Knowledge is limited and spread over measures of input, process, output and outcome</p> <p>Knowledge retained at a local level and power exercised by chief constables</p>	<p>Knowledge epitomized as an accounting for performance in terms of crime and detections</p> <p>Knowledge becomes centralized in iQuanta and power shifts to Home Office and Police Standards Unit</p>	<p>Knowledge shifts to encompass public satisfaction with policing and public fear of crime – unclear where this knowledge will be held, although seems likely to be retained centrally in iQuanta</p>

power and knowledge that Foucault (1975) saw as central to his arguments. In the first time period knowledge was diverse and local, with most power in the hands of chief constables. In the second period knowledge became focused on an accounting for crime and detections, with that knowledge centralized, giving the Home Office much more power than it had previously. In the third period knowledge expands to wider issues, and the renewed local focus may yet shift the balance of power. However, the centralization of knowledge remains undisturbed and the power of the Home Office is unlikely to be reduced.

Conclusions: performance measurement as knowledge and power

In policing, the triangle of sovereignty–discipline–government (Foucault, 1978 [1991]) has been shown to be a useful method by which to understand and explain the measurement and management of police performance. In particular, the research finds that, while there is no transition from sovereignty to discipline to government, there are shifts within each of these elements of Foucault’s triangle that explain the transition in how police performance is seen over different time periods.

Performance measures constitute knowledge about performance. As a managerial tool, performance measurement is used to identify areas for improvement. Performance measures are assessed against targets, in relation to trends over time, and as benchmark comparisons with “most similar” forces. Knowledge about performance is also a political tool. It provides a mechanism by which political parties are accountable to the electorate, and in turn use performance targets and measures as a means by which the concerns of the electorate are transmitted to those responsible for delivering public services.

Performance measures constitute what is accepted as knowledge, but the philosophical understanding of what knowledge is in turn constitutes performance measures. Hence, an understanding of what is to be measured, how it is to be measured, what targets are to be applied and how judgements about performance adequacy are made constitutes the performance measures. This involves the exercise of power.

Performance measurement as a key form of knowledge and power lies at the heart of the sovereignty that establishes government’s control over notionally “independent” chief constables. Sovereignty and disciplinary mechanisms can themselves be seen as governmental technologies. Discipline through hierarchical observation, normalizing judgement and the

examination is focused on performance measurement as the dominant form of knowledge about policing. Governmentality adds to discipline the dual technologies of the centralization of knowledge about performance and the power of intervention to improve performance from outside, which have led to the transition from a regime of performance measurement to one of performance management.

Knowledge is a form of power. The possession of knowledge not held by others gives a power to the holder to use that knowledge for particular purposes. This power tends to be held by those in government, typified by iQuanta and the PPAF. Nevertheless, processes for the acquisition, storage and sharing of knowledge also give power to those knowledge workers who can influence what is accepted as knowledge. The control of that which is constituted as knowledge and the processes by which that knowledge becomes constituted are elements of power that may fluctuate in the political arena between managers, civil servants and politicians.

Knowledge constitutes power and power constitutes knowledge, in the managerial, policy making and political dimensions. Performance measurement has been a tool by which the sovereignty of government over policing has been established, and performance measurement has been adopted as a disciplinary measure. These are important elements of the governmental technologies that use power to constitute knowledge deemed to be important for political purposes, which thereby reinforces the power of government over policing.

References

- Audit Commission (1998a). *Local Authority Performance Indicators 1996/97: Police Services*. London: HMSO.
- (1998b). *The Publication of Information Direction '98: Performance Indicators for the Financial Year 1999/2000*. London: HMSO.
- (1999). *Local Authority Performance Indicators 1997/98: Police Services*. London: HMSO.
- Bratton, W., with Knobler, P. (1998). *Turnaround: How America's Top Cop Reversed the Crime Epidemic*. New York: Random House.
- Carmona, S., and Gronlund, A. (2003). Measures vs actions: the balanced scorecard in Swedish law enforcement. *International Journal of Operations and Production Management*, 23, 1475–96.
- Carter, N., Klein, R., and Day, P. (1992). *How Organisations Measure Success: The Use of Performance Indicators in Government*. London: Routledge.
- CIPFA (2003). *Police Statistics: 2002–03 Actuals*. London: Chartered Institute of Public Finance and Accountancy.

- Clegg, S. R., Pitsis, T. S., Rura-Polley, T., and Marosszeky, M. (2002). Governmentality matters: designing an alliance culture for inter-organizational collaboration for managing projects. *Organization Studies*, 23, 317–37.
- Collier, P. M. (2001). The power of accounting: a field study of local financial management in a police force. *Management Accounting Research*, 12, 465–86.
- (2006a). Costing police services: the politicization of accounting. *Critical Perspectives on Accounting*, 17, 57–86.
- (2006b). In search of purpose and priorities: police performance indicators in England and Wales. *Public Money and Management*, 26(3), 165–72.
- English, J., and English, B. (1996). *Police Training Manual*. London: McGraw-Hill.
- Ezzamel, M. (1994). Organizational change and accounting: understanding the budgeting system in its organizational context. *Organization Studies*, 15, 213–40.
- Fitzgerald, M., Hough, M., Joseph, I., and Quereshi, T. (2002). *Policing for London*. Cullompton: Willan.
- Foucault, M. (1975). *Discipline and Punish: The Birth of the Prison* (trans. Sheridan, A.). Harmondsworth: Penguin.
- (1976 [1990]). *The History of Sexuality*, Vol. I, *The Will to Knowledge* (trans. Hurley, R.). Harmondsworth: Penguin.
- (1978 [1991]). Governmentality (trans. Braidotti, R.; rev. Gordon, C.). In Burchell, G., Gordon, C., and Miller, P. (eds.), *The Foucault Effect: Studies in Governmentality*. Chicago: University of Chicago Press.
- (1982). Afterword: the subject and power. In Dreyfus, H. L., and Rabinow, P. (eds.), *Michel Foucault: Beyond Structuralism and Hermeneutics*, 208–26. Hemel Hempstead: Harvester Wheatsheaf.
- Gordon, C. (1991). Governmental rationality: an introduction. In Burchell, G., Gordon, C., and Miller, P. (eds.), *The Foucault Effect: Studies in Governmentality*, 1–51. Chicago: University of Chicago Press.
- Gray, A., and Jenkins, B. (1995). From public administration to public management: reassessing a revolution? *Public Administration*, 73, 75–99.
- Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.
- Home Office (1993). *Police Reform: A Police Service for the Twenty-first Century: The Government's Proposals for the Police Service in England and Wales*, White Paper, Cm2281. London: HMSO.
- (1999). *Annual Report*. London: HMSO.
- (2001). *Policing a New Century: A Blueprint for Reform*, White Paper, Cm 5326. London: HMSO.
- (2004a). *Building Communities, Beating Crime: A Better Police Service for the 21st Century*. London: HMSO.
- (2004b). *National Policing Plan 2005–08: Safer, Stronger Communities*. London: HMSO.
- Hood, C. (1995). The “new public management” in the 1980s: variations on a theme. *Accounting, Organizations and Society*, 20, 93–109.
- Hoque, Z., Arends, S., and Alexander, R. (2004). Policing the police service: a case study of the rise of “new public management” within an Australian police service. *Accounting, Auditing and Accountability Journal*, 17, 59–84.

- Likierman, A. (1993). Performance indicators: 20 early lessons from managerial use. *Public Money and Management*, 13(4), 15–22.
- Maple, J., with Mitchell, C. (1999). *The Crime Fighter: How You can Make Your Community Crime Free*. New York: Broadway.
- Miller, P. (1990). On the interrelations between accounting and the state. *Accounting, Organizations and Society*, 15, 315–38.
- Miller, P., and Rose, N. (1990). Governing economic life. *Economy and Society*, 19, 1–31.
- ODPM (2004). *The Future of Local Government: Developing a 10-year Vision*. London: HMSO.
- OECD (1997). *In Search of Results: Performance Management Practices*. Paris: Organisation for Economic Co-operation and Development.
- Osborne, D., and Gaebler, T. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Harmondsworth: Plume.
- Public Services Productivity Panel (2000). *Improving Police Performance: A New Approach to Measuring Police Efficiency*. London: HM Treasury.
- Reiner, R. (2000). *The Politics of the Police*. 3rd edn. Oxford: Oxford University Press.
- Rogerson, P. (1995). Performance measurement and policing: police service or law enforcement agency? *Public Money and Management*, 15(4), 25–30.
- Rose, N., and Miller, P. (1992). Political power beyond the state: problematics of government. *British Journal of Sociology*, 43, 173–205.
- Smith, P. (1995a). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277–310.
- (1995b). Performance indicators and outcome in the public sector. *Public Money and Management*, 15(4), 13–16.

The development of composite indicators to measure health care performance

Rowena Jacobs, Peter Smith and Maria Goddard

Introduction

Over the past decade the measurement of comparative performance has become a dominant feature of health care systems across the world (Smith, 2002). Interest is apparent at every level, ranging from the performance of individual clinicians, providers and health plans, right up to the level of entire health systems. The ultimate rationale for collecting, analysing and publishing information on relative performance is to bring about performance improvement. However, the mechanism by which this works will vary, depending on the context within which health care systems operate. For example, market-orientated systems rely to a greater extent on the operation of competitive pressures from consumers, who may use comparative information when making choices about their health care insurers or providers (Porter and Teisberg, 2004). In public health systems, the publication of comparative performance information may form the basis for regulatory intervention.

It is widely acknowledged that health care performance is multi-dimensional. Policy makers and the public have a legitimate interest in a wide range of aspects of performance, such as efficiency, the quality of the health care process, accessibility, clinical outcomes and responsiveness (Institute of Medicine, 2001). There is now a plethora of information available for the measurement of relative performance, and interpreting such data is therefore becoming increasingly complex. One widely adopted approach to summarizing the information contained in disparate indicators of health care performance is to create a single composite measure. The rationale for developing such a composite measure is that no single indicator can hope to capture the complexity of system performance.

A composite indicator is an aggregated index comprising several individual performance indicators. Composite indicators are increasingly being used to measure the performance of, and also to rank, organizations and

institutions in economic, social and policy areas (Freudenberg, 2003). They integrate a large amount of information in a format that is easily understood, and are therefore a potentially valuable tool for conveying a summary assessment of performance in priority areas.

While the potential merits and disadvantages of composite as opposed to individual performance measures have long been acknowledged (Ridgway, 1956), the technical considerations in their development, and the impact these have on organizations, are still relatively poorly understood (European Commission, 2006). Their construction presents many methodological challenges, which, if not treated carefully and transparently, can leave them open to misinterpretation and potential manipulation. The accuracy, reliability and appropriateness of such indices needs to be explored if major policy, financial and social decisions hinge on the performance of organizations as measured by composite indicators.

The aim of this chapter is to examine the rationale underlying the development of composite measures of health care performance. We explore the advantages and disadvantages of constructing a composite indicator and describe the methodological choices made at each step in the construction of a composite. Steps include choosing the indicators to be used, transforming or standardizing the indicators, applying a system of weights to the indicators and then combining them to form the composite. The chapter also describes some examples of composite indicators in health care, highlighting good (and bad) practice in their development. We conclude with some thoughts on the future challenges in the development of composite performance indicators.

Arguments for and against the use of composite indicators

The arguments for developing composite indicators can be summarized as follows (Smith, 2002).

- (1) They place performance at the centre of the policy arena.
- (2) They can offer a rounded assessment of performance.
- (3) They enable judgements to be made on system efficiency.
- (4) They facilitate communication with the public and promote accountability.
- (5) They indicate which organizations represent beacons of best performance.
- (6) They indicate which organizations should be a priority for improvement efforts.

- (7) They present the “big picture” and can be easier to interpret than trying to find a trend in many separate indicators.
- (8) Depending on the construction of the composite, they can enable local policy makers to set their own priorities and seek out improvements along dimensions in which gains can be readily secured.
- (9) They may stimulate the search for better analytical methods and better-quality data.

On the other hand, composite indicators may lead to a number of dysfunctional consequences, and there are a number of arguments against their use (Smith, 2002).

- (1) By aggregating individual performance measures, composites may disguise serious failings in some parts of the system.
- (2) As measures become aggregated it becomes more difficult to determine the source of poor performance and where to focus remedial action.
- (3) The individual performance measures used in the composite are often contentious.
- (4) A composite that is comprehensive in coverage may have to rely on poor-quality data in some dimensions.
- (5) A composite that ignores some performance measures because they are difficult to measure may distort behaviour in undesirable ways.
- (6) The composite measure will depend crucially on the weights attached to each performance dimension. However, the methodology by which weights are elicited, and decisions on whose preferences they should reflect, are unlikely to be straightforward.

Steps to developing a composite indicator

There are a number of steps involved in constructing a composite indicator, each of which requires important judgements to be made by the analyst. At each step these judgements can, potentially, have a substantial impact on the ultimate outcome of the composite, and therefore they require careful consideration. The steps are as follows.

- (1) Choosing the entities to be assessed.
- (2) Choosing the organizational objectives to be encompassed in the composite.
- (3) Choosing the indicators to be included in the composite.
- (4) Transforming the measured performance of individual components.
- (5) Combining the components, using some sort of weighting.

- (6) Adjusting for environmental or other uncontrollable influences on performance.
 - (7) Adjusting for variations in expenditure (if a measure of efficiency is required).
 - (8) Conducting a sensitivity analysis to test the robustness of the outcome of the composite to the various methodological choices.
- These steps are examined in more detail below.

Choosing the units to be assessed and the organizational objectives to be encompassed

This is not a trivial matter, and hinges on the decision about what the boundaries of the units to be assessed are and what aspects of performance these units will be held responsible for. In health care, these boundaries are often somewhat blurred between primary care, secondary care, residential or long-term care and social services. Furthermore, the health care system extends well beyond health care, and attributing outcomes to individual sets of institutions is a difficult task.

These first two steps are clearly closely linked and hugely complex. Defining which institutions are to be held accountable for specific aspects of performance is operationally difficult. These are primarily political issues. We sidestep these and focus our discussion on the more technical aspects of the construction of composites in the following sections.

Choosing the indicators to be included

This step is probably one of the most important, and careful judgement is required. A different set of indicators will produce a different composite and hence a different set of performance rankings. The indicators included in the composite, the weight attached to them, and the reward schedule attached to the outcome will all have an effect on the effort that will be expended on trying to attain good performance on the included indicators, potentially at the expense of achieving on the excluded indicators.

In practice, composites are often either opportunistic and incomplete (measuring aspects of performance that are captured in existing data) or are based on highly questionable sources of data. Either weakness can cause serious damage to the credibility of the composite (Smith, 2002). The choice of indicators is most often constrained by data availability and may therefore give an unbalanced picture of health services. The excluded indicators may

be as important as (or more important than) the included indicators but simply more difficult to measure.

Aside from data availability, a number of other issues arise in the choice of performance indicators, namely the types of indicators to be included and the degree of collinearity between the indicators.

Regarding the first issue, there has been much debate about the pros and cons of different types of performance indicators in health care, in particular between the choice of *process* versus *outcome* measures. Health care is rarely demanded for its own sake. Rather, demand derives from the belief that health care will make a positive contribution to health status. For most patients and carers, health gain is the central indicator of the success of an intervention. A focus on outcomes directs attention towards the patient (rather than the services provided by the organization). Moreover, some widely accepted measures of health outcome (such as gains in quality-adjusted life years) are independent of the technologies used to deliver care, obviating the need for detailed scrutiny of the physical actions of the organizations.

One can seldom be confident, however, that outcome measures such as current health status are an indicator of current health system performance. For example, some health outcomes may take years to be realized, and it is clearly impractical to wait for them to emerge before coming to a judgement on performance. Furthermore, the collection of outcome data may impose high costs on the health system. Finally, there are issues around attribution and the extent to which health status can be attributed solely to the health care system. In such circumstances, it becomes necessary to rely on measures of health system process rather than health status outcome.

Outcome measures assess whether medical care has achieved certain results, while process measures assess whether medicine has been properly practised. Process measures relate to what is done for the patient and assess degrees of conformity to various “accepted standards” of care. Patients are becoming increasingly vocal in demanding that health care should be responsive to concerns over and above the health outcomes resulting from treatments. This concern with the “patient experience” covers issues as diverse as promptness, autonomy, empowerment, privacy and choice. Such performance measures may be appropriate particularly when there are large variations in the responsiveness of organizations, as indicated by hospital waiting times in many publicly funded health systems. In some circumstances, such as the management of chronic diseases, process measures will be far more relevant to patients than outcome measures (Crombie and Davies, 1998).

To reflect the breadth of performance in health care within a composite measure may, therefore, entail the inclusion of a wide spectrum of indicators between immediate process and eventual clinical outcome, with a variety of intermediate measures of outcome.

The second issue concerns the potential for performance indicators that are measuring similar aspects of performance to be highly correlated with each other. The concern is that, in the construction of a composite, the inclusion of variables that are highly collinear will effectively introduce some sort of double counting. It has therefore been argued that the inclusion of a reduced set of indicators based on a choice between those indicators with high correlations may be desirable for reasons such as parsimony and transparency.

Multivariate statistical methods are available to investigate the relationships between the indicators within a composite, including principal components analysis (PCA) and factor analysis (FA). These methods may be used to extract statistical correlations between indicators, enabling a core group of indicators to be identified that statistically best represents the remaining excluded indicators (Saisana and Tarantola, 2002). If statistical techniques are used to choose the variables for inclusion, then it is likely that highly collinear variables will be excluded through model specification tests for multicollinearity. The choice of one variable over an alternative, highly collinear variable may not alter rankings greatly, but it may affect the judgements on a small number of units with extraordinary performance in either of those dimensions. It may, therefore, be subject to dispute and challenge.

However, it is not strictly necessary from a technical point of view that highly collinear variables be excluded. For instance, if two perfectly collinear variables were included in the composite, then the particular dimension of performance that they measure will be included in the composite with a linear combination of the weights on the individual indicators. This is not problematic if the weights have been chosen correctly (Smith, 2002).

Transforming the indicators

There is no need for any transformation or standardization if it is possible to specify a weight that indicates the relative value to the composite of an extra unit of attainment in that dimension *at all levels of attainment* (Smith, 2002). In most cases, however, the indicators will be measured in different units that will reflect different weights in the composite, and therefore they will need to be transformed in some way before they can be aggregated in a

composite. Variables are transformed to a common basis to avoid the problems of mixing different units of measurement (such as labour, expenditure, events). Variables are also normalized to avoid problems with extreme values.

There are a number of methods available for transforming the underlying indicators, including ranking, normalizing, re-scaling, generating various types of ratio variables, logarithmic transformation and transforming variables to a categorical scale, all of which can impact on the final outcome of the composite indicator.

Combining the indicators

The different dimensions of performance measured on different scales (which are then transformed into a common scale) need to be combined in a meaningful way. This gives rise to some questions, namely what weights will be applied to the individual indicators, whose preferences will these weights reflect and how will they be elicited, and, finally, what rules will be applied to combine the transformed indicators into a composite?

When variables are aggregated into a composite they need to be weighted in some way. All variables may be given equal weight or they may be given different weights that reflect the priority, reliability or other characteristics of the underlying indicators (Freudenberg, 2003). As mentioned, weights may also be chosen to reflect a constant value for the relative price of the outputs (if variables are not transformed), although this may often be difficult to accomplish in practice. The relative weights indicate the amount of one objective that needs to be sacrificed in order to gain an extra unit of another objective.

Weights are, essentially, value judgements about the relative importance of different performance indicators and about the relative opportunity cost of achieving those performance measures. The rationale for the application of differential weights is that greater organizational effort will be used to achieve better performance on those dimensions that are considered more important. The weights that are attached to different performance indicators have a profound effect on the outcome of the composite index, and can change dramatically the ranking of a particular organization if an indicator on which the organization either excels or fails is given more (or less) weight.

The weights can be obtained either from statistical techniques or by using the judgement of individuals based on survey methods to elicit

their preferences. There are, of course, fundamental considerations around whose preferences will be used in the application of those weights, whether it be the preferences of policy makers, providers, purchasers, patients or the public. The weights used will usually reflect a single set of preferences, whilst evidence suggests that there exist a great diversity in preferences across policy makers, individual providers and the broader public. This should temper the notion of presenting the composite as “objective” (Smith, 2002).

In addition to important considerations around whose preferences are elicited, there are also a variety of approaches for eliciting preferences or values. These include single and multiple voting, simple scoring and scaling, budget allocation (Appleby and Mulligan, 2000), public opinion polls and direct interview techniques (Shaw *et al.*, 2001), as well as more complicated methods such as analytic hierarchy process and conjoint analysis, each with its own advantages and disadvantages (Mullen and Spurgeon, 2000). There appears to be little consensus as to which the preferred technique is (Dolan *et al.*, 1996) and it is likely that the different methods will lead to the emergence of different preference sets.

In contrast to using a process of attaching weights to the transformed indicators, the indicators can be combined into a composite indicator by applying a set of decision rules. An example of this is in the construction of the scorecard for acute hospitals in the star rating system in England, by which a complicated algorithm with a set of sequential decision rules is applied to determine the ultimate star rating (composite indicator). In this case, the preference set of policy makers is used and an implicit set of weights is generated through the algorithm.

One reason for applying decision rules is to try to ensure that certain minimum requirements are met before organizations become eligible for any further reward. Applying these decision rules to attain minimum standards may be particularly pertinent with the accreditation process of hospitals. Evidence suggests that the application of such rules, and subtle changes to their application, can be hugely influential in the final outcome of the composite measure (Jacobs, Smith and Goddard, 2005).

Adjusting for environmental or uncontrollable factors

Some organizations have to operate in more adverse environmental circumstances, which may make the attainment of performance outcomes more difficult for them. It is argued, therefore, that adjustments might be

made to take into account these exogenous environmental conditions when measuring their performance.

There may be many causes of variation and exogenous influences on performance (Jacobs and Dawson, 2002; Gravelle and Smith, 2002), including:

- differences in the health status of the population (for example, age and gender mix, co-morbidities, case mix);
- the external environment (for example, geography);
- differences in the resources used (or an inappropriate mix of resources);
- differences in the quality of services provided;
- variations in institutional arrangements (for example, specific hospital features);
- different priorities regarding objectives;
- different accounting treatments (and data conventions);
- data errors;
- random fluctuations (some measures may be more vulnerable to fluctuations beyond the control of the system); and
- differences in the effectiveness of the system in achieving the chosen objectives – the key issue of interest.

Thus, composite indicators of health care performance may not be measuring just the performance of health care agencies but also population and environmental characteristics, such as unemployment, education and housing conditions, that influence health status and the use of health services. In trying to isolate the differences in the effectiveness of organizations in meeting the chosen objectives, all the other influences need to be taken into account. It may not always be possible, or policy-relevant, to correct for all these sources of variation under all circumstances.

One element in the debate concerns the delineation of the boundaries of the health care system under investigation. If the composite is focused on the performance of the whole health care “system” (all the activities of government) then there is less argument for controlling for exogenous factors. This wider view argues that, though public health care agencies may be unable to influence certain factors such as unemployment, these factors can be influenced by other parts of government (Appleby and Mulligan, 2000). However, if the interest is in a more narrow definition of health care agencies, then it may be very important to undertake the adjustments.

Another issue relates to the degree to which the allocation of resources makes it unnecessary to adjust for exogenous factors. It could be argued that, if the health care funding formula used to distribute funds to health care

organizations is designed to enable all health care agencies to produce equal levels of performance, given their different environmental circumstances, then there may be no need to adjust for exogenous circumstances in the construction of the composite (Smith, Rice and Carr-Hill, 2001). These formulae take into account population characteristics, and it can be argued that an indirect form of standardization is therefore being carried out via the funding system. This assumes that organizations are adequately compensated for their differences in health needs. Thus, if the funding formula has been correctly designed to permit equal performance, and all relevant aspects of performance can be captured by the composite, then there would be no need to control for exogenous factors. Variations in performance need be examined then only as the difference between observed outcomes of different organizations. The degree to which this is a reasonable argument depends on views about the accuracy and comprehensiveness of the formulae used to allocate resources.

If an adjustment for uncontrollable influences on performance is to be made, it may be made either at the final stages after having constructed the composite indicator and/or prior to this, at the level of the individual performance indicators. When adjustment is made at the level of the composite, there may be technical difficulties in trying to incorporate exogenous influences on performance. In general there are two approaches in the literature to modelling exogenous influences on a set of performance scores. The first is to incorporate them in a one-step model in which they are included as an input in the production process. The second is a two-step model in which the model is first set up to exclude exogenous factors and then, in a second step, the performance scores are explained by exogenous influences. While the two-step approach may be more transparent and practical, it is often contentious which factors are considered truly exogenous and should be left out of the first step. In addition, the variables in the first and second stages are likely to be highly correlated, leading to biased results (Simar and Wilson, 2002). The essential point is that there is no generally accepted method for taking into account environmental variables at the level of the composite scores, or for testing whether an environmental variable has a significant influence on the organization and the resultant performance of the unit.

Alternatively (or additionally), individual indicators are often adjusted for differences in the health status of the population before they are aggregated. Risk adjustments are made to the data to include the age, sex, type of admission, length of stay and co-morbidity profiles of the relevant population.

However, there are technical difficulties with this approach as well. Alternative methods of risk adjustment usually give rise to different results, and may lead to large variations in measured performance (Iezzoni *et al.*, 1995, 1996).

Examining variations in efficiency

As a final step in the construction of a composite indicator, regulators may be interested in exploring the efficiency with which organizations use resources in relation to achieving the performance measured on the composite. This leads to the examination of performance (outputs or outcomes) in relation to resource use or the costs devoted to the attainment of the performance (inputs) – which is typically a measure of efficiency (Smith, 2002).

If organizations are to be ranked against their composite scores, then they should be given budgets that in some sense give them equal opportunities to secure equal composite scores if they are technically and allocatively efficient. If organizations in adverse environments are generally scoring poorly on a composite measure relative to their less disadvantaged counterparts, this may be because they are consistently less efficient. Taking account of variations in efficiency is one way of testing whether this is the case. On the other hand, as mentioned, it may be because they are not being funded well enough to secure higher scores. If this is the case, the funding formula needs to skew resources further towards organizations in adverse circumstances in order to offer them a level playing field (Hauck, Shaw and Smith, 2002; Smith, 2003).

Sensitivity analysis of the construction of the composite indicator

As illustrated by each of the preceding steps in the construction of a composite indicator, there are a variety of difficulties that can arise with respect to selecting, transforming, weighting and aggregating variables into a composite. The outcomes and rankings of individual units on the composite may depend crucially on the judgements made and the decisions taken at each step. Sensitivity analysis is a useful tool with which to explore the robustness of rankings to the inclusion and exclusion of certain variables, changes in the weighting system, using different transformation methods and setting different decision rules to construct the composite (Freudenberg, 2003). While most analysts would probably agree that sensitivity analysis is considered good practice, in reality this is seldom exercised.

International examples of the development and use of composite indicators

In this section we describe the development and use of six published composite indicators in health care to illustrate some of the challenges in the construction of each. To varying degrees, all these composite indicators have been appraised for their potential strengths and weaknesses. All these examples require careful consideration at each step in their methodology and they therefore share many commonalities. However, we highlight the steps in their construction that are particularly pertinent and that raise particular challenges. These are listed in table 18.1, along with a summary of their key features.

United States Medicare

In 2000 and 2003 Jencks and colleagues produced a series of twenty-two quality indicators of the care delivered to Medicare beneficiaries and constructed state-level composite indices from these indicators. They examined these state-level indicators for the periods of 1998/9 (baseline) and 2000/2001 (follow-up) and examined changes in performance across the range of indicators. The quality indicators were abstracted from state-wide medical records for inpatient fee-for-service care (sixteen indicators) and from Medicare beneficiary surveys or Medicare claims for outpatient care (six indicators).

Indicators were chosen for six clinical areas: acute myocardial infarction (AMI – six indicators), heart failure (two indicators), stroke (three indicators), pneumonia (seven indicators), breast cancer (one indicator) and diabetes (three indicators). The choice of indicators tended to over-represent inpatient and preventive services and under-represent ambulatory care and interventional procedures. The indicators were also not risk-adjusted, and hence focused on process measures rather than outcomes.

In the first study (Jencks *et al.*, 2000), each of the states and non-state territories is ranked on each of the measures, and so the percentage score is transformed to an ordinal scale ranging from 1 to 52. A composite performance measure was produced by computing each state's average rank. In the follow-up study (Jencks, Huff and Cuerdon, 2003), absolute improvement was also calculated on each indicator and relative improvement, or the reduction in error rate. To summarize the overall changes in performance at

Table 18.1: Summary of international examples of the use of composite indicators and key considerations

Setting	Units covered	Period covered	Performance indicators	Key features	Issues/considerations
United States Medicare	52 states	2000 and 2003	22 indicators, 6 areas: AMI (6); heart failure (2); stroke (3); pneumonia (7); breast cancer (1); diabetes (3)	Calculate average rank in 2000 and absolute and relative improvement	Choice of indicators; weighting of indicators
Canada	63 regions	annual	15 indicators, 6 areas: outcomes (2); prenatal care (3); community health (2); elderly services (2); efficiency (3); resources (3)	Indicators re-scaled and weights used based on “expert opinion”	Weighting of indicators; examining efficiency
United Kingdom	Health authorities	2000	6 indicators	Attach weights based on public preferences using different methods of eliciting preferences	Choice of indicators; transforming the indicators
WHO	191 countries	2000	5 indicators	Scores transformed and summed using weights based on mainly WHO staff views	Construction of indicators; transforming the indicators; weighting of indicators; examining efficiency

Table 18.1: (cont.)

Setting	Units covered	Period covered	Performance indicators	Key features	Issues/considerations
England	NHS providers	2000/2001 – 2005/6	Around 35 indicators, 4 areas: key targets; clinical focus; patient focus; capacity and capability	Transform indicators into categorical variables and apply sequential decision algorithm	Choice of indicators; weighting of indicators; adjusting for environmental factors
United States Medicare and Medicaid	278 hospitals	2003–2006	34 indicators, 5 areas: AMI (9); CABG (8); heart failure (4); pneumonia (7); hip and knee replacements (6)	Compute 5 composites with weights according to number of process and outcome measures in each composite	Choice of indicators; transforming the indicators; weighting of indicators

the state level, they calculated the median amount of absolute and relative improvement across the set of indicators in the state. They also calculated the rank of each state on each quality indicator based on the 2000/2001 performance, and the rank on each quality indicator based on relative improvement. They then calculated the average rank for each state across the twenty-two indicators, and league-tabled them according to their average rank based on 2000/2001 performance as well as on their relative improvement. They found that a state's average rank on the twenty-two indicators was highly stable over time with a correlation of 0.93 between the two periods. The better-performing states appeared to be concentrated geographically in the northern and less populated regions (for both periods), while the geographic patterns of relative improvement by state were more patchy. The report showed overall improvement across twenty of the twenty-two indicators.

There were some concerns in connection with the choice of indicators (process measures), which the authors acknowledged, and in connection with the reliability of the data used to construct some of the indicators (Smith, 2002). Furthermore, since all the indicators were given equal weight, the use of six indicators for AMI would give that clinical area a higher contribution in the composite than, say, breast cancer, for which there was only one indicator.

The use of the league table ranking as the basis for the composite also implicitly assumes that identical differences in ranking are equally important, regardless of where in the league table they occur. The incentive is, therefore, for states to concentrate on activities in which they can more readily secure a movement up the league table, rather than those that offer the most potential health gain.

Canadian regional health care

Macleans magazine (MM) is a major mass-circulation Canadian magazine that publishes an annual Health Report, in which it ranks Canadian regions according to their health care. Composite performance scores ranged from 89.5 in North/West Vancouver, British Columbia, to 73.4 in North Bay/Huntsville, Ontario, in 2001.

The scores are produced using data published by the Canadian Institute for Health Information on a series of annual reports as well as a series of health indicators for the sixty-three largest regions (covering 90 per cent of the population) (Canadian Institute for Health Information, 2001a, 2001b).

The *Macleans* report uses fifteen performance indicators grouped into six categories.

- (1) Outcomes: (i) life expectancy at birth; (ii) heart attack survival.
- (2) Prenatal care: (i) proportion of low-birthweight babies under 2,500 g; (ii) percentage of babies born by Caesarean section; (iii) percentage of vaginal births after Caesarean section.
- (3) Community health: (i) hip fractures; (ii) pneumonia and flu hospitalization of persons over sixty-four.
- (4) Elderly services: (i) hip replacements; (ii) knee replacements.
- (5) Efficiency: (i) possible outpatients – hospitalizations for conditions not requiring admission; (ii) early discharge – variation from expected length of stay; (iii) preventable admissions – hospitalizations for conditions considered preventable by appropriate ambulatory care.
- (6) Resources: (i) physicians per 100,000; (ii) specialists per 100,000; (iii) local services – percentage of hospitalizations generated by local residents.

MM re-scaled each of the fifteen indicators to have a mean of 80 and a standard deviation of 10 (with a higher score implying better performance). Within each of the six categories, the scores on the performance indicators were combined using weights “based on expert opinion”. The six categories were then combined using the following weights: outcomes 0.2; prenatal care 0.2; community health 0.2; elderly services 0.1; efficiencies 0.2; resources 0.1. This sequential approach to assigning weights, first to performance indicators within categories and then to categories, allows a more careful treatment of priorities, albeit that the weighting scheme is very rudimentary and the preferences of the “experts” may not necessarily reflect those of the public. Moreover, the inclusion of efficiency indices leads efficiency to be treated not as the extent to which objectives are secured in relation to expenditure but, rather, as simply another objective that contributes to the concept of performance (Smith, 2002).

British health authorities

In 2000 the UK television broadcaster Channel 4 commissioned researchers at the King’s Fund to explore the public’s relative preferences or health care priorities (Appleby and Mulligan, 2000). They produced a ranked list of English and Welsh health authorities and Scottish health boards according to a composite indicator based on selected aspects of performance, designed to reflect the relative weight attached by the public to these measures of National Health Service (NHS) performance. Researchers were limited on the number of indicators with which they could feasibly survey the public,

and thus they restricted their choice to six indicators chosen from readily available data.

- (1) The number of deaths from cancer (per 100,000).
- (2) The number of deaths from heart disease (per 100,000).
- (3) The number of people on hospital waiting lists (per 1,000).
- (4) The percentage of people on waiting lists for more than twelve months.
- (5) The number of hip operations (per 100,000).
- (6) The number of deaths from “avoidable” diseases (tuberculosis, asthma, etc., for which there are effective clinical interventions that would prevent death) (per 100,000).

The focus of the study was to attach weights to each indicator based on public preferences. A polling organization surveyed 2,000 people across England, Scotland and Wales to obtain their preferences. Three methods were used for eliciting preferences: ranking from most- to least-desired indicator; “budget pie”, in which respondents were asked to allocate a “budget” of sixty chips between the six performance indicators; and conjoint analysis, asking respondents to choose between different mixes of options. The weights for each indicator were taken from the budget pie method.

The researchers were concerned that some of the raw performance indicators had skewed distributions and were not all measured on the same scale, and data were standardized to a Z score. The rankings generated from these transformed and standardized indicators differed from the original ranks, with the average change in ranking being fourteen places and a 0.81 rank correlation.

Again, some concerns may be raised over the choice of indicators and their coverage, as well as over the sensitivity of the rankings of the health authorities arising from the methods of transforming the data.

The World Health Report 2000

The composite index of health system performance produced for 191 countries by the World Health Organization (WHO) in *The World Health Report 2000* has been the subject of much debate (WHO, 2000; Williams, 2001; Smith, 2002; Appleby and Street, 2001; Nord, 2002). The index was based on a weighted sum of attainment across three broad areas: health, financial fairness and responsiveness. Five dimensions were captured in total, as the health and responsiveness indicators considered both the overall level of attainment and their distribution:

- (1) overall health outcomes;
- (2) inequality in health;

- (3) fairness of financing;
- (4) overall health system responsiveness; and
- (5) inequality in health system responsiveness.

The first dimension – average population health – was measured in terms of disability-adjusted life expectancy (DALE). The second dimension – equity of health – was measured primarily in terms of equality in survival for the first five years of life. Fairness in financing was measured by creating an index ranging from 0 to 1, defined as the ratio between total expenditure on health and total non-food expenditure. Responsiveness was meant to capture how well the health care system responded to basic non-health expectations in the population, and was expressed as an index covering seven aspects of responsiveness (the dignity and autonomy of patients, freedom of choice of provider, the quality of amenities, prompt attention, confidentiality and access to social support networks). Each dimension of responsiveness was scored by around 2,000 “key informants” from thirty-five countries, who answered questions about their own country and were then asked to give a score for responsiveness as a whole.

The seven aspects were then ranked in order of importance by 1,000 people and weights were assigned based on the rankings. Mean scores on each aspect were multiplied by the weights and summed to give an overall responsiveness score. The final dimension – equity in responsiveness – was calculated by asking informants to make judgements about the subgroups they thought were treated with less responsiveness than others. Scores were assigned to subgroups based on the number of times the country informants mentioned them, multiplied by the share of that group in the population. The products were summed and transformed to give an overall score. Finally, the scores on each of the five dimensions were transformed to a 0 to 100 scale and summed using weights of either 0.25 or 0.125, based on the views of about 1,000 people from 123 countries, half of whom were WHO staff.

A second index was also created, using the WHO composite measure of attainment to estimate overall health system efficiency. In this second stage of econometric modelling an attempt was made to capture relative performance by looking at the difference between what was actually achieved in each country (attainment) and the maximum it could be expected to achieve given its resources. The latter was measured by health system expenditure and exogenous influences (human capital captured by years of education).

The debate about the appropriateness of the WHO rankings has been widespread and has touched on most of the potential disadvantages

concerning composite indices listed earlier. Some of these relate to the political context in which WHO operates, the ideological values underpinning the production of the rankings, their purpose and the lack of transparency in the WHO methods (Navarro, 2000, 2001, 2002; Williams, 2001; Almeida *et al.*, 2001).

The major methodological concerns about the creation of the composite measure have revolved around the nature of the underlying data; the transformation of the data; the weighting system; and how the weights were elicited. A large number of technical and analytical judgements were made in the WHO rankings, and, whilst many have argued about the potentially dysfunctional consequences of poor-quality analysis and resultant inappropriate policy responses, the publicity and debate around the WHO exercise have put system performance and composite performance indicators much higher on the agenda of policy makers and analysts.

The UK star rating system for NHS providers

In September 2001 the first set of performance star ratings was published by the Department of Health for acute NHS trusts in 2000/2001 (Department of Health, 2001). The star ratings are a composite index score given to each NHS organization and place them into one of four categories, from three stars (the highest levels of performance) to no stars (the poorest levels of performance). In July 2002 the second set of star ratings was published by the Department of Health, now covering acute trusts, specialist trusts, ambulance trusts and indicative ratings for mental health trusts for 2001/2 (Department of Health, 2002). Primary care trusts (PCTs) received a separate publication, describing their performance against a range of suitable indicators, but not a rating. In July 2003 all NHS providers were covered and the Commission for Health Improvement (CHI), the independent regulator of NHS performance, took over responsibility for performance ratings from the Department of Health (CHI, 2003). In July 2004 and 2005 two more sets of ratings were published by the Healthcare Commission (previously the CHI; Healthcare Commission, 2004, 2005).

The methodology for the star ratings has remained relatively constant, although some important changes have been made to the individual indicators covered. Broadly, the trust star ratings comprise four areas: key government targets; clinical focus; patient focus; and capacity and capability. The key government targets are the most significant factors in determining overall performance ratings. The broader range of indicators make up a “balanced

scorecard” to refine the judgement on ratings. Performance against key targets is assessed in terms of whether the target has been achieved, whether there has been some degree of underachievement or whether the target was significantly underachieved (threshold-type variables).

In general terms, the methodology entails transforming the underlying key targets and performance indicators into categorical variables of either three or five categories. The performance indicators in the patient, clinical, and capacity and capability focus areas are categorized into one of five performance bands, with 5 points being awarded for the best performance and 1 for the worst. The thresholds for deciding the cut-offs are not necessarily the same for each variable. Individual band scores are combined to produce an overall score per area. All the indicators are equally weighted within their scorecard area in such a way as to ensure that, despite the differing numbers of indicators, each scorecard area carries the same weight. The clinical indicators are published with confidence intervals, which means that performance is split into three bands depending on whether the organization’s confidence interval overlaps the England average for the indicator.

A complex six-step process is then imposed whereby a sequential set of decisions on achievement on the various key variables determines the ultimate star rating outcome.

The star rating outcome has a significant reward and penalty schedule attached to it, since trusts that obtain a three-star rating for a consecutive number of years may apply for foundation status, which will give them significant financial and managerial decision-making freedoms and autonomy from central involvement. The incentives for ensuring a good outcome on the star rating are therefore very strong.

Both the construction and the impact of star ratings have been questioned. Methodological concerns have tended to focus on such issues as the partial coverage of health system objectives, the degree to which external factors influence achievements, and the relative weight given to aspects of the ratings (Kmietowicz, 2003; Cutler, 2002; Snelling, 2003; Miller, 2002).

Pay for performance for Medicare and Medicaid services

In July 2003 Premier Inc., a nationwide organization of not-for-profit hospitals in the United States, and the Centers for Medicare and Medicaid Services (CMS) launched the Hospital Quality Incentive (HQI) Demonstration Project (Premier Inc., 2005; CMS, 2004). It is a three-year programme to

establish if financial incentives can improve the quality of inpatient care. CMS will reward participating hospitals that achieve superior performance by increasing their payment for Medicare patients. The project covers five clinical areas:

- (1) acute myocardial infarction (nine indicators);
- (2) coronary artery bypass graft (CABG) (eight indicators);
- (3) heart failure (four indicators);
- (4) community-acquired pneumonia (seven indicators); and
- (5) hip and knee replacement (six indicators).

Hospital performance for each of the five categories is aggregated into a composite score to establish baseline performance. Continuous variables are converted to rate-based measures by establishing a threshold and then calculating the number of patients that received care within the established limits. Each composite consists of a process score (twenty-seven indicators are process) and outcome score component (seven indicators are outcome), with proportional weights applied according to the number of each type of indicator in the category. The composite *process* score in each category is created by summing the numerator and denominator values for each indicator and then dividing the totals. The composite *outcome* score in each category is created by generating a survival index of actual divided by expected survival rate. Each is then multiplied by the component weighting factor. The choice of indicators and the weighting process is therefore likely to have an important impact on the composite score.

The composite score is then used to identify the hospitals that perform in the top two deciles that will be eligible for incentive payments. Those in the top decile of quality for a given clinical area will be given a 2 per cent bonus of their Medicare payments for the given condition, while hospitals in the second decile will be paid a 1 per cent bonus. Composite quality scores are calculated annually. In year three, those hospitals that do not achieve performance improvements above the baseline will have adjusted payments. Hospitals will receive a 1 per cent lower payment for a clinical condition if they score below the ninth decile baseline and 2 per cent less payment if they score below the tenth decile baseline level for year one.

A total of 278 hospitals were included in the project, which is voluntary, and after one year around 270 hospitals improved their performance significantly. With a voluntary scheme, of course, self-selection of better-performing hospitals in the five clinical areas is probably a major issue, and the system may have limited impact for those hospitals performing very poorly.

Conclusions

We have shown that the construction of composite indicators is not straightforward, and many methodological issues need to be addressed carefully if the results are not to be misinterpreted and manipulated. Although the construction of composite measures raises many detailed technical issues, the impact of judgements and decisions made in the context of these methodological challenges is substantial, and they have important ramifications that extend beyond technicalities. As the final two examples show, when the outcome of the composite score is used for regulatory purposes and for payment the construction of the composite becomes crucial. The use and publication of composite performance measures can generate both positive and negative behavioural responses, and if significant policy decisions rest on the outcome of the composite (such as foundation status for hospitals in the United Kingdom's NHS, or financial reward in the United States) it is important to have a clear understanding of the potential risks involved in constructing a composite and arriving at such a ranking.

As yet there are many unresolved issues and challenges in constructing composite indicators, particularly in connection with transforming and weighting the underlying indicators and adjusting for environmental factors. Evidence suggests that different ways of treating each of these steps in the process can result in very different rankings for health care organizations (Jacobs, Smith and Goddard, 2005). While composite indicators and league tables enjoy a high profile in the popular press, the danger is that premature publication can damage health care organizations, but the hope is that it can also lead to heightened research activity and a more satisfactory methodology.

From a policy point of view, the conceptual limits of composite indicators should be borne in mind and their publication should be accompanied by explanations of the choice of indicators, the transformation method and weighting structure. Explanations of the limits of the composite may help with its interpretation, and also with making the process more transparent so that it can be clear what policy objectives are being maximized.

References

- Almeida, C., Braveman, P., Gold, M. R., Szwarcwald, C. L., Ribeiro, J. M., Miglionico, A., Millar, J. S., Porto, S., Costa, N. R., Rubio, V. O., Segall, M., Starfield, B., Travessos, C.,

- Uga, A., Valente, J., and Viacava, F. (2001). Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet*, 357, 1692–7.
- Appleby, J., and Mulligan, J. (2000). *How Well is the NHS Performing? A Composite Performance Indicator Based on Public Consultation*. London: King's Fund.
- Appleby, J., and Street, A. (2001). Health system goals: life, death and . . . football. *Journal of Health Services Research*, 6(4), 220–5.
- Canadian Institute for Health Information (2001a). *Health Care in Canada 2001: A Second Annual Report*. Ottawa: Canadian Institute for Health Information.
- (2001b). *Health Indicators 2001*. Ottawa: Canadian Institute for Health Information.
- CHI (2003). NHS performance ratings: acute trusts, specialist trusts, ambulance trusts 2002/2003. London: Commission for Health Improvement, www.chi.nhs.uk/ratings/.
- CMS (2004). *CMS HQI Demonstration Project: Composite Quality Score Methodology Overview*. Baltimore: Centers for Medicare and Medicaid Services.
- Crombie, I.K., and Davies, H.T.O. (1998). Beyond health outcomes: the advantages of measuring process. *Journal of Evaluation in Clinical Practice*, 4(1), 31–8.
- Cutler, T. (2002). Star or black hole? *Community Care*, 30 May, 40–1.
- Department of Health (2001). *NHS Performance Ratings: Acute Trusts 2000/01*. London: Department of Health.
- (2002). *NHS Performance Ratings and Indicators: Acute Trusts, Specialist Trusts, Ambulance Trusts, Mental Health Trusts 2001/02*. London: Department of Health.
- Dolan, P., Gudex, C., Kind, P., and Williams, A. (1996). Valuing health states: a comparison of methods. *Journal of Health Economics*, 15, 209–31.
- European Commission (2006). Composite indicators: an information server on composite indicators. Ispra, Italy: European Commission, Joint Research Centre, http://farmweb.jrc.ec.europa.eu/ci/JRC_2006EPI_PR.htm.
- Freudenberg, M. (2003). *Composite Indicators of Country Performance: A Critical Assessment*, Science, Technology and Industry Working Paper 2003/16. Paris: Organisation for Economic Co-operation and Development.
- Gravelle, H., and Smith, P.C. (2002). Sources of variations in measured performance. In Smith, P.C. (ed.), *Some Principles of Performance Measurement and Performance Improvement*, report prepared for the Commission for Health Improvement, 3–5. York: University of York, Centre for Health Economics.
- Hauck, K., Shaw, R., and Smith, P.C. (2002). Reducing avoidable inequalities in health: a new criterion for setting health care capitation payments. *Health Economics*, 11(8), 667–77.
- Healthcare Commission (2004). 2004 performance ratings. London: Healthcare Commission, <http://ratings2004.healthcarecommission.org.uk/>.
- (2005). 2005 performance ratings. London: Healthcare Commission, <http://ratings2005.healthcarecommission.org.uk/>.
- Iezzoni, L.I., Schwartz, M., Ash, A.S., Hughes, J.S., Daley, J., and Mackiernan, Y.D. (1995). Using severity-adjusted stroke mortality rates to judge hospitals. *International Journal for Quality in Health Care*, 7(2), 81–94.
- (1996). Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*, 34(1), 11–28.

- Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: Institute of Medicine of the National Academies, Committee on Quality of Health Care in America.
- Jacobs, R., and Dawson, D. (2003). Variation in unit costs of hospitals in the English NHS. *Journal of Health Services Research and Policy*, 8(4), 202–8.
- Jacobs, R., Smith, P. C., and Goddard, M. (2005). How robust are hospital ranks based on composite performance measures? *Medical Care*, 43(12), 1177–84.
- Jencks, S. F., Cuerdon, T., Burwen, D. R., Fleming, B., Houck, P. M., Kussmaul, A. E., Nilasena, D. S., Ordin, D. L., and Arday, D. R. (2000). Quality of medical care delivered to Medicare beneficiaries: a profile at state and national levels. *Journal of the American Medical Association*, 284(13), 1670–6.
- Jencks, S. F., Huff, E. D., and Cuerdon, T. (2003). Change in the quality of care delivered to Medicare beneficiaries, 1998–1999 to 2000–2001. *Journal of the American Medical Association*, 289(3), 305–12.
- Kmietowicz, Z. (2003). Star rating system fails to reduce variation. *British Medical Journal*, 327, 184.
- Miller, N. (2002). Missing the target. *Community Care*, 21 November, 36–8.
- Mullen, P., and Spurgeon, P. (2000). *Priority Setting and the Public*. Abingdon: Radcliffe Medical Press.
- Navarro, V. (2000). Assessment of the World Health Report. *Lancet*, 356, 1598–601.
- (2001). World Health Report 2000: response to Murray and Frenk. *Lancet*, 357, 1701–2.
- (2002). The World Health Report 2000: can health care systems be compared using a single measure of performance? *American Journal of Public Health*, 92(1), 31–4.
- Nord, E. (2002). Measures of goal attainment and performance: a brief, critical consumer guide. *Health Policy*, 59(3), 183–91.
- Porter, M. E., and Teisberg, E. O. (2004). Redefining competition in healthcare. *Harvard Business Review*, 82(3), 65–76.
- Premier Inc. (2005). HQI demonstration overview. Washington, DC: Premier Inc., www.premierinc.com/all/quality/hqi/index.jsp.
- Ridgway, V. F. (1956). Dysfunctional consequences of performance measurements. *Administrative Science Quarterly*, 1(2), 240–7.
- Saisana, M., and Tarantola, S. (2002). *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*. Ispra, Italy: European Commission, Joint Research Centre.
- Shaw, R., Dolan, P., Tsuchiya, A., Williams, A., Smith, P. C., and Burrows, R. (2001). *Development of a Questionnaire to Elicit Preferences regarding Health Inequalities*, Occasional Paper no. 40. York: University of York, Centre for Health Economics.
- Simar, L., and Wilson, P. W. (2002). *Estimation and Inference in Two-stage Semi-parametric Models of Production Processes*. Louvain: Université Catholique de Louvain, Institut de Statistique.
- Smith, P. C. (2002). Developing composite indicators for assessing health system efficiency. In Smith, P. C. (ed.), *Measuring Up: Improving the Performance of Health Systems in OECD Countries*, 322–46. Paris: Organisation for Economic Co-operation and Development.

- (2003). Formula funding of public services: an economic analysis. *Oxford Review of Economic Policy*, 19(2), 301–22.
- Smith, P. C., Rice, N., and Carr-Hill, R. (2001). Capitation funding in the public sector. *Journal of the Royal Statistical Society, Series A*, 164(2), 217–57.
- Snelling, I. (2003). Do star ratings really reflect hospital performance? *Journal of Health Organization and Management*, 17, 210–23.
- WHO (2000). *The World Health Report 2000: Health Systems: Improving Performance*. Geneva: World Health Organization.
- Williams, A. (2001). Science or marketing at WHO? *Health Economics*, 10, 93–100.

Perversity in public service performance measurement

Mike Pidd

Introduction: government, targets and measurement

The last twenty years have seen the use of performance measurement and performance management mushroom in government departments and agencies in the United Kingdom and elsewhere. A review of the use of performance measurement in various countries can be found in *Public Management Reform: A Comparative Analysis*, by Pollitt and Bouckaert (2004). Performance measurement is not a new feature of public administration, but it has become much more prominent in the United Kingdom, the United States, New Zealand and some other countries during the last two decades. A positive review of these developments in the United Kingdom was produced by the government's Comptroller and Auditor General in 2001 (National Audit Office, 2001). A rather more critical review can be found in the report produced by the Public Administration Select Committee in 2003 following its extensive discussions (Public Administration Select Committee, 2003).

There are many different aspects of the public sector that could be measured. These include issues of productivity, traditionally the domain of economists, which was the major concern of the Atkinson Review (Office of National Statistics, 2005), set up by the National Statistician "to advance methodologies for the measurement of government output, productivity and associated price indices in the context of the National Accounts . . .". These macro-level concerns, which stemmed from the previous habit of economists to assume that the appropriate measure of output for government activities was the input cost (which means that there can never be measured productivity gains), are not the concern of this chapter. Rather, this contribution focuses on the use of performance measurement in the provision of public services to citizens as part of the contract between governors and governed.

There are several basic assumptions underpinning the arguments running through the material of this chapter. The first is that measurement is an essential part of management, which is an issue that will be considered further in the next section. The second is that people will measure things anyway, so it makes sense to encourage them to do this properly. The third argument, which will also be unpacked in this chapter, is that people seem reluctant to accept that the negative effects of performance measurement must be included in any evaluation of its effects.

Why measure public sector performance?

Why is the measurement of public sector performance so important when discussing the provision of public services? At around the same time as the Atkinson Review was announced, the Royal Statistical Society (RSS) completed a review that recommended best practices for performance measurement in public services (Bird *et al.*, 2003). This RSS review suggested that there are three important reasons for measuring the performance of UK public services.

To see what works

The first reason cited in Bird *et al.* (2003) relates to the need for policy to stand on a firm evidence base. That is, performance measurement is needed *to see what works*. This recognizes that there are many competing voices advocating different approaches to provision and that these options need to be compared. If there are several options available the performance of each can be measured, and this information can be used for comparison to determine the best way to proceed. Alternatively, the impact of a single policy might be measured by comparing its costs and benefits to see if it is worthwhile. This form of measurement, which is essentially quantitative in nature, has many attractions as long as it is done properly. It requires the careful use of statistical methods and designed comparisons; some of the possible approaches are described in Boyne *et al.* (2002).

This form of measurement can be regarded as contributing to a learning loop in which people learn about what works and take appropriate action for improvement. Done properly, this learning loop corresponds to Deming's PDCA (plan–do–check–act) cycle, which is a generic model of continuous improvement for total quality management – see figure 19.1. Indeed, measurement is fundamental to this view of quality improvement. However, it is

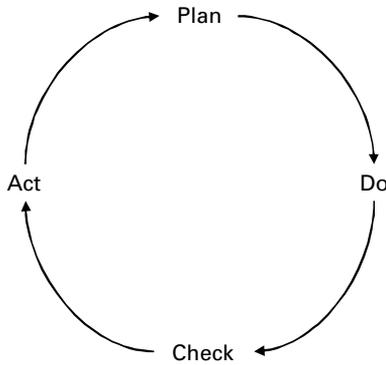


Figure 19.1: Deming's PDCA cycle

important to consider *who* learns in this learning cycle, since this will depend on how the measurement is implemented. If it is implemented and managed by those who deliver the public services in question, then this allows them to learn and to improve. The measurement may be regarded positively if it allows service providers to make improvements based on agreed measures of performance. However, if the measurement is imposed by some central authority across a range of service providers with little regard for local sensitivity, it seems likely that there will be resentment and a risk of game playing.

To identify competences

The second reason for measuring public sector performance suggested by Bird *et al.* (2003) is *to identify competences*. This is a form of post hoc audit that carries the connotation that performance will be compared, either with other providers or with some norm, so as to identify good and bad performers. Hence, the focus is less on what people do (as in measurement to find out what works) and more on measuring how people have performed – that is, not just “What?” and “How”, but also “Who”. Needless to say, this may be unwelcome, and it needs to be carried out fairly and carefully. Also writing for a statistical audience, Goldstein and Spiegelhalter (1996) strongly criticize the way in which published performance indicators are sometimes unthinkingly used to construct league tables. As in most sports, mid-table teams are all pretty much of a muchness, and if the range of values produced by the metrics is narrow there may be no real difference between any of those being ranked in a league table.

It is important to realize that performance indicators are, in essence, based on simple statistical models, and that any resulting performance estimates should carry reasonable estimates of error. Once these are

included in the measures then much of the apparent ranking in league tables disappears, since there is no statistically significant difference in performance between many of the units ranked in the tables. Hence, rewarding people because of their high ranking in these league tables or punishing them because they seem to do badly is a dangerous game unless there are very real differences. Measurement with this aim in mind is usually enforced from outside and may be carried out by dedicated auditors who are independent of particular service providers. Needless to say, if people's careers and salaries depend on the outcome of this measurement there is a great risk of game playing, as well as danger that it may stymie the risky experiments and innovations that are needed to produce positive change in an organization.

To support public accountability

Thirdly, Bird *et al.* suggest that measurement is needed *to support public accountability*. Public services are funded mainly through taxation and it seems reasonable that citizens should know whether their taxes are being spent wisely. For example, the United Kingdom's Labour government headed by Tony Blair massively increased expenditure on the National Health Service. It seems reasonable to ask whether this has resulted in improvements and whether these changes represent value for money. However, though public accountability is important, providing appropriate measurement is more difficult. One main problem is that the public, however construed, do not seem very interested beyond the brief flurry of excitement that follows, say, the annual publication of star ratings for hospitals and other health care providers. These appeared in newspapers and were discussed on television and the radio for a day or two, but they were quickly forgotten by the public, and there is no evidence that service users or their relatives use the information to choose a service provider when a choice is available (Marshall *et al.*, 2003). Investigating a different area of public service, Miller (2003) reports that local authorities find it difficult to engage in meaningful public consultation as required by national legislation. The public, it seems, is awkward enough not to behave like the public.

There are two reasons why this may be so. The first is the difficulty of defining what we mean by the "public". Discussing this issue in relation to Canadian health care, Contandriopoulos, Denis and Langley (2004) identify three different "publics" whose support is often solicited by politicians. The

first is the reified public, which is often spoken of as the “real public”, the “average man or woman” or the “ordinary citizen”. This abstract group is defined as having no involvement in the health care system, and therefore being relatively ignorant about the health care system. It has “the same status as the abstract citizen upon whom the democratic system rests” (Contandriopoulos, Denis and Langley, 2004, 1580). Anyone who gets involved in the system will lose their status as “average man or woman”.

The second group is named “the public’s voices”, and includes those who participate in decision-making processes, such as citizens’ representatives, community organizations and citizens’ groups. These are people who project themselves as effective public representatives and who are often seen by others as partial and with conflicts of interest. They are engaged in a symbolic struggle for the right to speak for the “public” and may attempt to present themselves as “average men or women”. Finally, there are the users, patients, families and others who consume services and who have opinions and interests that differ significantly from those of the “general public”. Which group is the public who will act as customers for this fourth root definition? Which is the public that has a right to know and what do they have a right to know? How can performance data be provided that is meaningful, and is it to be aimed at the uninformed reified public, the public’s voices or service users?

This brings us to the second reason for this lack of interest: the difficulty of providing information in forms appropriate to the various publics. Providing data as star ratings provides useful headlines, but service users are well aware that things are more complicated. In the case of health care, it seems that they are more likely to make choices on the basis of personal information and advice from those whom they trust (Steele, 2003). That is, the data are presented in terms that are not meaningful to people or that are not convincing. This may turn out to be an insuperable problem, since presentations such as star ratings were introduced, partially at least, to make things comprehensible (Hibbard *et al.*, 2002). Research in the United States (Hibbard and Peters, 2003) shows that improved presentation can increase public interest, though still only slightly. Making measures relevant may require very careful data presentation, but even then it is unclear whether this will increase public interest.

To summarize the discussion so far, Bird *et al.* (2003), representing the Royal Statistical Society, a technical body, argue that there are three reasons for performance measurement in the provision of public services: to see what works, to identify competences and to support public accountability.

All three are very sensible, but the list is incomplete, and there are other, very important, reasons why performance measurement is used in this sector.

Missing reason one: to support central control

The most obvious reason for performance measurement missing from the analysis in Bird *et al.* stems from the need for central government to exercise control over the services for which it is responsible, and on which it spends tax revenues. Recent years have seen the rise of what is known as new public management, of which there are many definitions. Most agree that it emphasizes decentralized service provision through a variety of agencies, rather than through monolithic directly governed bureaucracies. With this in mind, Hood (1991, 4, 5), one of the earliest commentators on the popularity of NPM, lists eight doctrinal components of NPM, which include

- explicit standards and measures of performance; and
- greater emphasis on output controls.

Putting these two doctrines together suggests that a major reason for performance measurement in the provision of public services stems from the need to retain control when service provision is decentralized through agencies or subcontracted to the private sector. That is, to the list proposed by Bird *et al.*, we must add that performance measurement is used *to enable central control*.

“Cybernetics” is defined as the science of control, and the term itself has its roots in the Greek word usually transliterated as *kybernetes* (steersman) and the Latin word *gubernator* (governor). Control by the feedback of information is the central concept of cybernetics, and it is this that lies at the heart of performance measurement to enable central control. The concept of cybernetic control is shown in figure 19.2, in which the output of some process is monitored and compared to some target. Adjustments are made to the system inputs and to the organization of the system itself, so as to steer the outputs towards the targets. The idea is simple enough, and finds widespread use in the physical devices that we use in our daily lives. Since some people regard management accounting as focusing on organizational control, it is not surprising that cybernetic control has been widely discussed in its literature.

A significant contribution to the debate about the use of cybernetic control is made by Hofstede (1981), who argues that the conventional model

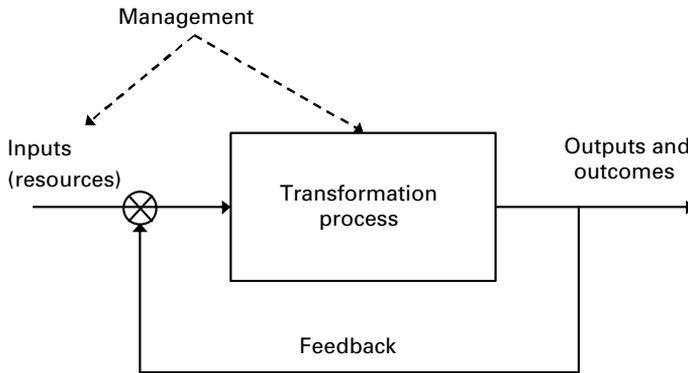


Figure 19.2: The cybernetic control metaphor

of control is based on the simple, mechanistic metaphor from cybernetics, shown in figure 19.2, in which information is fed back to a comparator and appropriate corrective action is taken. Hofstede argues that such an approach works in controlling an activity only if it meets all four of the following conditions.

- (1) The objectives of the activity are unambiguous; thus there is no disagreement amongst stakeholders, no uncertainty about means–ends relationships and no environmental turbulence.
- (2) The outputs from the activity are quantifiable.
- (3) There is perfect knowledge about how to intervene (what action to take) and what the effects of that action will be.
- (4) The activity is repetitive so that there is the opportunity to learn.

It should be clear that, though some activities in the provision of public services do meet all four requirements, many do not. Since the political realm is concerned with the gaining and exercise of power, it is highly likely that there will be disagreement about means–ends relationships. Because many public projects run for long periods and are subject to changes in political power, there is often considerable environmental turbulence. In addition, it is often unclear how to measure the outcomes of a public service and whether an observed outcome is a result of a control intervention or has been caused by other factors. In this vein, it is normal to distinguish between outputs and outcomes. Sometimes both can be measured without too much difficulty; for example, in the health sector, an outcome would be infant mortality and a related output might be the number of infants immunized against particular infections. However, even in this case there are other reasons why infant mortality may change, such as poverty, poor housing and malnutrition, and these may confound the effects of immunization.

Table 19.1: Hofstede on control

Appropriate type of control	Objectives	Measurability	Knowledge	Repetition
Routine	Unambiguous	Total	Complete	Frequent
Expert	Unambiguous	Total	Complete	Never or rare
Trial and error	Unambiguous	Total	Limited or none	Frequent
Intuitive	Unambiguous	Total	Limited or none	Never or rare
Judgemental	Unambiguous	Limited or none	Doesn't matter	Doesn't matter
Political	Ambiguous	Doesn't matter	Doesn't matter	Doesn't matter

There is little doubt that activities that meet Hofstede's four criteria can be controlled through systems based on the cybernetic metaphor. Many systems in the public sector cannot be so controlled, however, because they do not meet these requirements, and table 19.1 shows other types of control that have been identified by Hofstede as appropriate when all four criteria are not satisfied. Despite this insight, it is still tempting to assume that this problem can be resolved by changing the nature of these activities so that they do meet these criteria.

Noordegraaf and Abma (2003) discuss the place of ambiguity and uncertainty in the provision and control of public services. They argue that both ambiguity and uncertainty are fundamental to some aspects of the public sector and that to imagine otherwise is a serious mistake. Therefore, and perhaps less obviously, attempts to redesign work to make it less uncertain and less ambiguous may be unwise. Figure 19.3 summarizes the implications of Noordegraaf and Abma's insight. It shows a spectrum that, for simplicity, shows only the extreme points and is based on the organization of the UK public sector. The left-hand side of figure 19.3, labelled with examples from the day-to-day provision of benefits and the routine collection of taxes, clearly meets Hofstede's requirements for routine, cybernetic control. There is very little uncertainty and ambiguity, since those employed in these roles are required to follow strict protocols, based on legislation or the law, that define their actions.

The right-hand end of figure 19.3, however, is labelled with examples of health care and education in which those who are engaged in the activity (the "delivery agents") are highly skilled professionals and have great discretion in how they act. Schön (1983) argues that professionals are reflective practitioners who learn from encounters with their clients and who develop their skills in so doing. Early in their careers most professionals are

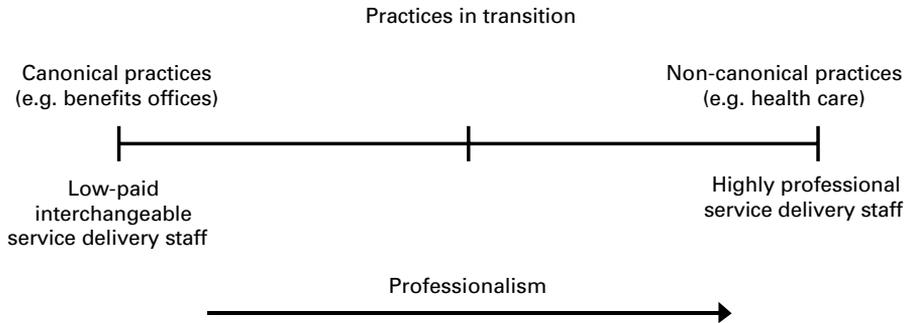


Figure 19.3: Ambiguity, uncertainty and non-canonical practices

uncertain, and they cope with this by adhering fairly strictly to protocols and are usually closely supervised as they do so. Mature professionals, however, operate somewhat differently, and learn to cope with ambiguity and uncertainty, though not by pretending that it does not exist. It may, therefore, be a very serious mistake to treat activity conducted by professionals in the same way as that carried out by less skilled staff. This mistake may be evident in two ways. First, when a control system simply ignores the ambiguity and uncertainty. This will lead to great dissatisfaction and, likely as not, to game playing, since many of these professionals are smart people. The second occurs when well-intentioned attempts are made to deskill professional activities by attempting to reduce the freedom to learn and develop new responses. This serves as a major disincentive to innovation – hardly a desirable outcome.

The point being made about professionals is not intended as a defence of privilege, nor as a class-ridden statement that close control is acceptable for interchangeable, clerical drones but not for the articulate middle class. It is based on the observation of Noordegraaf and Abma that there are activities and programmes in the public sector in which ambiguity and uncertainty are fundamental and cannot be designed out. In these circumstances, Hofstede (1981) argues that other modes of control, as shown in table 19.1, are much more appropriate.

Missing reason two: symbolic action

The other important reason for performance measurement in the public sector comes from the nature of political activity. To gain and keep power in a democracy, politicians must be seen to engage in activity that reflects their

contract with the electorate. Hence, there are very good reasons why politicians – and, as a consequence, public servants – may engage in performance measurement *for symbolic reasons*.

This justification for performance measurement is very different from the previous four reasons, though the other four may be called into play as a smokescreen for political activity. There is little point in considering whether symbolic action is a good or bad thing; it simply is true that it occurs, and it seems to be an essential part of the political world: it is important for political actors to be seen to act, to do something. If this action resonates with the zeitgeist, in this case NPM, then so much the better, and this is another reason why performance measurement is currently a popular theme in the public sector of many countries.

Understanding perversity and dysfunctionality in performance measurement

Having discussed the main reasons why performance measurement is pervasive in the public sectors of some countries, it is now time to consider what can go wrong and why. This is the main theme of this chapter, not with the aim of gloating over other people's problems or pointing fingers at the mistakes of others but, rather, to support learning so that the practice of performance measurement may be improved. It should also be noted that performance measurement is an idea that is applied at many different levels in any organization, including those in the public sector. At one extreme, it can form part of appraisal schemes for individual employees, and, at the other, balanced scorecards are in common use for entire public sector organizations. In this chapter the focus is on the use of performance measurement in assessing outputs and outcomes from activities and programmes, not for assessing the performance of individuals in their work.

That performance measurement may be dysfunctional is hardly a new insight. The first volume of *Administrative Science Quarterly* included a landmark article by Ridgway (1956) discussing this topic, though not specifically in the context of public sector organizations. From time to time others have returned to this topic, and one of the most widely cited is an article by Smith (1995), which appeared about forty years after Ridgway's but is specifically concerned with the public sector, and its examples are set in the United Kingdom. The aim of this section is to consider some of the reasons

why this perversity and dysfunctionality occur, based mainly on the previous section, which considered the reasons why performance measurement is used in organizations that provide public services.

Clarity of vision

The earlier sections of this chapter have discussed five different reasons why performance measurement is used and has become widespread in the provision of public services. It should, then, be obvious that those introducing and maintaining such schemes need to be clear about their objectives in so doing. It is, obviously, possible to employ performance measurement successfully for more than one of the five reasons discussed above, but even in these circumstances some clarity of vision is essential.

One approach to developing this clarity of vision is to consider, in abstract terms, the main features of a system that implements performance measurement for each of these five reasons. Soft systems methodology (SSM), discussed by Checkland (1999) and Checkland and Scholes (1999), employs the idea of a *root definition* to conceptualize an abstract representation known as a human activity system. This is a purposive (i.e. serving a purpose) system that expresses some purposeful (i.e. deliberate or willed) human activity. In other words, a human activity system is a conceptual model of some purpose-designed entity in which humans engage in deliberate activity. A root definition is usually presented using the CATWOE acronym, developed from the initial letters of the following six terms, as discussed by Pidd (2003, 125ff.).

- (1) *Customers*. These are the immediate beneficiaries or victims of what the system does. It can be an individual, several people, a group or groups. This is very close to the total quality management notion that the customer is the next person to receive the work in progress. It indicates what happens to the output from the system and forms part of the external relations of the system.
- (2) *Actors*. In any human activity system there are people who carry out one or more of the activities in the system; these are the actors. They form part of the internal relations of the system. There may be several actors or several groups, and their relationships also form part of the internal relations of the system.
- (3) *Transformation process*. This is the core of the human activity system, in which some definite input is converted into some output and then

passed on to the customers. The actors take part in this transformation process. The process is an activity and its description therefore requires the use of verbs. Ideally, a root definition should focus on a single transformation.

- (4) *Weltanschauung*. This is the, often taken for granted, outlook or world view that makes sense of the root definition being developed. It is important to specify this because any system definitions can make sense only with some defined context. Thus a root definition needs only a single *Weltanschauung*.
- (5) *Ownership*. This is the individual or group responsible for the proposed system, in the sense that it has the power to modify it or even to close it down. This can overlap with the actors of the system.
- (6) *Environmental constraints*. All human activity systems operate within some constraints imposed by their external environment. These might be, for example, legal, physical or ethical. They form part of the external relations of the system and need to be distinguished from its ownership.

Pidd (2005) provides a detailed description of a set of root definitions that embody the five reasons for measuring performance. Table 19.2 summarizes root definitions for the five reasons for measurement performance discussed earlier. There is very little difference in the actors, owners and environmental constraints cited in the definitions, but there are important differences in the customers, the transformations and the *Weltanschauungen*. If we assume that all five root definitions regard the actors as being operated by specialists on behalf of the government and policy staff within constraints set by available finance, skills and the law, the root definitions can be written as follows.

When performance measurement is introduced to see what works, this is done on behalf of policy makers, who wish to see evidence on which policy can be based, in the belief that policy should be developed from a sound evidence base. This allows the policy staff, if not the service providers, to learn, and, it is hoped, this will lead to improved policy. When performance measurement is introduced to identify competences, this is done on behalf of both government and policy staff, who require data to enable comparison and do so in the belief that this provides a useful basis on which people and organizations can be praised or blamed. This is clearly linked to the idea of central control, to which we shall return shortly. When performance measurement is introduced to support public accountability, this is done on behalf of the public (however defined),

Table 19.2: Root definitions

	To see what works	To identify competences	To support public accountability	To enable central control	For symbolic reasons
C	Policy staff	Government and policy staff	The public	Government and policy staff	Government
A	Specialists	Specialists	Specialists	Specialists	Specialists
T	Provide evidence for policy development	Provide data to allow comparison of performance	Provide reliable data for the public	Maintain control of decentralized services	Be seen to do something
W	Good policy is based on sound evidence	Need to do this to allow praise and blame	The public has a right to know	Important to maintain control	Important to be seen to do something in accordance with zeitgeist
O	Government and policy staff	Government and policy staff	Government and policy staff	Government and policy staff	Government
E	Law, finance and skills	Law, finance and skills	Law, finance, skills and public interest	Law, finance and skills	Law, finance and skills

who need performance information and who have a right to this in a democratic society. In other words, the customers, transformation and Weltanschauung are very different from the previous two. This indicates that it may be very unwise to assume that the same measurement system can be used.

What of performance measurement introduced to maintain central control? Table 19.2 shows that it is justified as serving the needs of government and policy staff, who wish to retain central control of decentralized services, since they will not allow full autonomy to the service providers. As discussed above, this is done by introducing targets against which progress and performance are assessed. When linked to performance measurement to establish competences, praise and blame may be employed to encourage compliance. Finally, when performance measurement is introduced for symbolic reasons, this is done by the government (taken to represent the political community), which wishes to be seen to be doing something, preferably something that chimes with the zeitgeist, since this is an important part of the symbolic struggle within the political domain.

It is sometimes assumed that performance measurement will always support learning, but this is unlikely to be so for everyone involved. If performance measurement is introduced centrally to see what works then central planners may learn, but local provider units will not automatically do so unless included in the processes. Put another way, performance measurement to see what works should be designed so that the service providers are part of the learning process. Some form of learning will certainly occur if performance measurement is introduced to enable central control or to establish competences, but this may be narrow, disinclined to innovation and fearful. However, it should be clear that using performance measurement mainly for reasons of public accountability or as part of a symbolic struggle will require very different system designs from those needed for the other three. It should also be clear that neither is concerned with learning, at least not in any conventional sense, nor with control against targets.

Each of the five cited reasons for performance measurement provides a rationale that can be defended. When designing and implementing a system that aims to measure aspects of the performance of public service provision, it is important to be clear about the reasons for its introduction. Without this, the introduction will be much less successful than hoped. The five reasons cited here provide a sensible starting point for a discussion that might underpin such planning.

Reducing dysfunctionality

Even when a performance measurement system is introduced for well-thought-out reasons there remains the strong probability of unexpected side effects and dysfunctional behaviour. Smith (1995) suggests eight different types of dysfunctional behaviour that have appeared when quantitative performance indicators (PIs) have been introduced in public service provision and made public. Smith is not arguing that using and publishing PIs is a bad thing or that they should be avoided. Rather, he argues that they are sometimes introduced and published in somewhat clumsy ways, which leads to the unintended consequences that he identifies and defines as follows.

- (1) *Tunnel vision* “can be defined as an emphasis by management on phenomena that are quantified in the performance measurement”. The goals of public service organizations are rarely one-dimensional and this means that any measurement must embody multiple criteria. Some are easier to measure than others, however, hence the oft-repeated warning that *if it can be counted, it will count*.
- (2) *Suboptimization* “is the pursuit of narrow local objectives by managers, at the expense of the objectives of the organization as a whole”. It is seen in most bureaucracies, which, by their hierarchical nature, encourage people to focus on their own local concerns. Doing so, however, may degrade the higher-level performance of the organization.
- (3) *Myopia* is “the pursuit of short term targets at the expense of legitimate long term objectives”. This is a problem in many public services for which outcomes are apparent only a considerable time after action has been taken. To keep people focused, short-term targets are sometimes introduced, and these will inevitably loom large in people’s minds.
- (4) *Measure fixation* occurs because outcomes are difficult to measure, which leads to a natural tendency to use performance indicators based on measurable outputs. However, outputs (e.g. number of patients treated) are not the same as outcomes (e.g. morbidity from a particular cause). Measure fixation “can be defined as an emphasis on measures of success rather than the underlying objective”.
- (5) *Misrepresentation* “is the deliberate manipulation of data so that reported behaviour differs from actual behaviour”. In other words, it is a form of fraud in which data that form parts of PIs are distorted

Table 19.3: Dysfunctionalities and reasons for measurement

	To see what works	To identify competences	To support public accountability	To enable central control	For symbolic reasons
Tunnel vision	✓			✓	
Suboptimization				✓	
Myopia		✓		✓	
Measure fixation	✓	✓		✓	✓
Misrepresentation		✓		✓	✓
Misinterpretation	✓		✓	✓	
Gaming		✓		✓	
Ossification				✓	

and manipulated so as to create a good impression with some target audience.

- (6) *Misinterpretation* occurs because public service organizations are typically large and complex, and therefore understanding their performance is not straightforward. “Thus even if the available data were a perfect representation of reality, the problem of interpreting the signals emerging from the data is often extremely complex.” Such misinterpretation was a major reason for the creation of the RSS’s working party (Bird *et al.*, 2003).
- (7) *Gaming* “can be defined as the deliberate manipulation of behaviour to secure strategic advantage”. It is seen, for example, when managers negotiate budgets that are very relaxed or targets that are easy to achieve. If next year’s targets are to be based on this year’s performance, this provides a perverse incentive to perform only moderately this year so as to avoid stretching targets next year.
- (8) *Ossification* is “organizational paralysis brought about by an excessively rigid system of performance evaluation”. It can happen when a performance indicator is past its sell-by date and has lost its purpose, but no one can be bothered to revise or remove it. The performance measurement system can provide an incentive for managers not to innovate, for fear of damaging their measured performance.

Table 19.3 takes the five reasons for introducing performance measurement and links these to the unintended consequence identified in Smith (1995). Since Smith is mainly concerned with the publication of PIs intended for internal control purposes, it should be no surprise that all eight consequences

are possible if measurement is used to enable central control. It is, of course, possible that any of the eight could occur whatever the original vision for the performance measurement; some unintended consequences are more likely, however, for particular forms of measurement.

When measurement is introduced *to see what works* then the greatest risks seem to be tunnel vision, with its measure fixation and misinterpretation. Tunnel vision has a focus on aspects that are easier to measure and that therefore may form a large part of any judgement about what works. Measure fixation is a possible consequence of tunnel vision, in which the larger objectives disappear from view. Finally, there is misinterpretation, in which the customers, government and policy staff misread the signals that are provided by the PIs and wrongly draw conclusions about success and failure.

When measurement is introduced *to identify competences*, there is a clear danger that those whose performance is being measured will see the likelihood of praise and blame and so will operate accordingly. Hence, there is a great risk of myopia, measure fixation, misrepresentation and gaming by those charged with the provision of a public service, since their livelihood may depend on it. In the early days of star ratings in the United Kingdom's NHS, one measure of performance included was the delay before a patient was seen by a nurse after arriving in accident and emergency. Trusts were quick to employ so-called "hello nurses", whose job was to greet the patient, thus meeting the target but doing nothing whatsoever for the health of the patient. In addition, as with measurement to see what works, there is always the risk that the customers for this measurement – government and policy staff – may misinterpret the performance data that they receive.

When measurement is used *to support public accountability* the main risk seems to be of misinterpretation, and this is particularly acute when performance data are summarized in league tables and star ratings. The latter are intended to simplify the interpretation of performance, but they can obscure the real meaning. For example, the star ratings for NHS acute hospital trusts requires them to hit targets related to waiting times and financial performance, but they say nothing about the clinical performance of the trust. There is the risk that potential patients and their families may think otherwise. League tables are also dangerous, though they are popular with newspapers. As an example, in 2008 the research performance of UK university departments will be measured in a research assessment exercise(RAE).

All departments will be told what proportion of their research falls into each of five categories:

4*	World-leading
3*	Internationally excellent
2*	Recognized internationally
1*	Recognized nationally
Unclassified	Below 1*

Leaving aside the problem of defining what terms such as “world-leading” actually mean, a department may be told, for example, that 10 per cent of its work is 4*, 20 per cent is 3*, 30 per cent is 2*, 30 per cent is 1* and 10 per cent is unclassified. Once the results are published the league tables will appear, and each university will start to put its own particular spin on the results. A profile with five parts is not specific enough to use in a league table and so most commentators will compute an average of some kind. Hence, for our hypothetical university department, its weighted average score is $(4 \times 0.1) + (3 \times 0.2) + (2 \times 0.3) + (1 \times 0.3) = 1.6$. League tables could be constructed on this basis, but this ignores how many staff were included, since departments can choose to omit staff members. If the department in question submitted twenty staff its volume score could be computed as $20 \times 1.6 = 32$. League tables could be constructed on this basis. If the newspaper knew how many staff were not included, then another score could be computed. Suppose the department has thirty staff and submitted the research of only twenty for assessment; a suitable ratio then might be the ratio of the actual volume score to the maximum volume score. This would be $32 / (4 \times 30) = 32 / 120 = 0.27$, if each person were required to submit four research outputs for assessment.

Several such league tables are possible, and each table will have different rankings, but there is a hidden computation that provides further scope for confusion. The rules for the RAE insist that the proportion of work in each of the five grades should be rounded to the nearest 5 per cent. Though the rules prescribe a particular way of performing this rounding, its effect could be perverse, and there could be no real difference between departments that receive different scores – the known problem of mid-table performance in league tables discussed earlier.

If measurement is introduced for symbolic reasons, to produce totems that can be worshipped, then two particular risks come to the fore: measure fixation and misrepresentation. Both relate to the deliberate or unwitting abuse of statistical measures, a topic well covered by authors such as Boyle (2001), Huff (1991) and Brignell (2000).¹ Whether we like it or not, politicians are wont to quote whatever statistics suit them in trying to make a particular case, regardless of whether they have been produced by independent bodies. Deliberate or not, the effect is to induce measure fixation in those who believe what they hear, which could itself be based on misinterpretation. As an example, the published statistics show that waiting times for inpatient treatment in acute hospitals in the United Kingdom have been significantly reduced since they became part of star ratings. This provides useful numerical support for current government policies. It is unclear, however, what has happened to other aspects of the health delivery system that are not included in star ratings. It is possible that squeezing the system in one place leads to distortions elsewhere – it may not, of course, but we simply do not know.

Dysfunctionalities and cybernetic control

Since all the dysfunctionalities discussed by Smith (1995) figure in measurement to enable central control, it seems worth considering this in a little more depth. This is not intended as an argument that no central control is needed; far from it. It is, however, a call for clarity when discussing performance measurement. Put simply, in some circumstances the use of quantitative PIs as performance measures for central control may make things worse. If control is needed it may be better exercised in other, more appropriate, ways.

As summarized earlier, Hofstede (1981) discusses the different forms of control that are needed for different types of activity, and argues that cybernetic control, with its emphasis on unambiguous measurement, typically expressed in quantitative PIs, is inappropriate in some circumstances. It is effective when objectives are unambiguous, when accurate and appropriate measurement is possible, when all concerned have adequate knowledge of what is happening and when the activity is repeated many times. Clearly, this

¹ See also Brignell's website: www.numberwatch.co.uk.

fits some low-level activities very well and can be made to fit some higher-level ones, but it brings with it great dangers.

These are discussed by Mintzberg, whose *Harvard Business Review* article “Managing government, governing management” (Mintzberg, 1996) provides a useful summary of his argument. In essence, he argues that the cybernetic model is part of “Capital-letter Management”, by which he means a parody of truly effective business management. Capital-letter management, he argues,

finds its full realization in the performance control model, the motto of which could be Isolate, Assign and Measure. [...] The overall organization is split into “businesses” that are assigned performance targets for which their managers are accountable. So the superstructure plans and controls, whilst the microstructure executes. All very tidy. But not necessarily very effective. [...] [T]he performance model decentralizes in order to centralize, it loosens up in order to tighten up. And tightening up comes at the expense of flexibility, creativity and individual initiative. (Mintzberg, 1996, 80, 81)

Sometimes the aim of central control is to ensure that protocols are rigidly followed, in which case the cybernetic model is fine. However, there are many circumstances in which it is not.

Focusing on for-profit organizations, in chapter 21 Rob Austin and Jody Hoffer Gittell examine anomalous high performance – which occurs when the following basic principles of performance measurement, stemming from the usual cybernetic metaphor, are deliberately flouted.

- (1) Performance should be clearly and unambiguously defined, preferably before the event. That is, those whose performance is being measured should be clear about the aspects of performance that matter. Since, in some cases, performance has to be defined after the event, this is possible only if there is strong reciprocal trust.
- (2) Performance should be measured accurately.
- (3) Rewards should be contingent on performance.

As before, these seem very sensible principles, but they imply that the primary aim of performance management is to induce compliance, which is related to a view that employees will not perform well unless there is some extrinsic motivation. In this view, motivation is provided by the market contract or by participation in a bureaucracy. The idea is to align workers’ preferences with those of the organization.

However, Austin and Gittell provide a series of vignettes from anomalous high-performing organizations that break some or all of these rules.

The vignettes suggest that a different approach is possible: one that encourages workers to develop their own intrinsic motivation, which can be used to help them align their goals to those of the organization. In a similar vein, in chapter 22 Thomas Ahrens and Chris Chapman describe a high-performing restaurant chain, which also broke Austin and Gittell's three principles listed above. It did so by loosely coupling specific performance metrics with specific managerial responses – that is, local restaurants were allowed to take different actions in response to local conditions and circumstances. In this way, flexible and appropriate responses were encouraged, rather than slavish adherence to a central rule book. Ahrens and Chapman argue for such loosely coupled performance management and measurement, especially for organizations that operate in diverse contexts, whether in the public or private sector. The idea is that performance management systems should encourage flexible and appropriate responses using local knowledge.

High performance, it seems, can be achieved by exploiting intrinsic motivation and relying on people to develop their own flexible responses based on local conditions. This does not, of course, mean that there is a free-for-all in which public sector employees can do whatever they wish. However, they can be given autonomy and encouraged to develop their own performance metrics. Rejecting bureaucracy need not mean embracing cybernetic control, and this is very important, since, as Noordegraaf and Abma (2003) point out, ambiguity and uncertainty are fundamental to many activities in the public sector. These characteristics are not best treated by pretending that they do not exist.

References

- Bird, S. M., Cox, D., Harvey, G., Tim, H., and Smith, P. C. (2003). *Performance Indicators: Good, Bad, and Ugly*, report of the Royal Statistical Society Working Party on performance monitoring in the public services. London: Royal Statistical Society.
- Boyle, D. (2001). *The Tyranny of Numbers: Why Counting can't Make Us Happy*. London: Flamingo.
- Boyne, G., Farrell, C., Law, J., Powell, M., and Walker, R. (2002). *Evaluating Public Management Reforms: Principles and Practice*. Milton Keynes: Open University Press.
- Brignell, J. (2000). *Sorry, Wrong Number!: The Abuse of Measurement*. London: Brignell Associates.
- Checkland, P. B. (1999). *Systems Thinking, Systems Practice: Includes a 30-year Retrospective*. Chichester: John Wiley.

- Checkland, P. B., and Scholes, J. (1999). *Soft Systems Methodology in Action*. Chichester: John Wiley.
- Contandriopoulos, D., Denis, J.-L., and Langley, A. (2004). Defining the “public” in a public healthcare system. *Human Relations*, 57(12), 1573–96.
- Goldstein, H., and Spiegelhalter, D. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). *Journal of the Royal Statistical Society, Series A*, 159, 385–443.
- Hibbard, J. H., and Peters, E. (2003). Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. *Annual Review of Public Health*, 24, 413–33.
- Hibbard, J. H., Slovic, P., Peters, E., and Finucane, M. L. (2002). Strategies for reporting health plan performance information to consumers: evidence from controlled studies. *Health Services Research*, 37, 291–313.
- Hofstede, G. (1981). Management control of public and not for profit activities. *Accounting, Organisations and Society*, 6(3), 193–211.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19.
- Huff, D. (1991). *How to Lie with Statistics*. Harmondsworth: Penguin.
- Marshall, M. N., Shekelle, P. G., Davies, H., and Smith, P. C. (2003). Public reporting on quality in the United States and the United Kingdom. *Health Affairs*, 22(3), 134–45.
- Miller, D. (2003). The virtual moment. *Royal Anthropological Institute*, 9(1), 57–75.
- Mintzberg, H. (1996). Managing government, governing management. *Harvard Business Review*, 74(3), 75–83.
- National Audit Office (2001). *Measuring the Performance of Government Departments: Report by the Comptroller and Auditor General*, HC 301. London: HMSO.
- Noordegraaf, M., and Abma, T. (2003). Management by measurement? Public management practices amidst ambiguity. *Public Administration*, 81(4), 853–73.
- Office of National Statistics (2005). *Atkinson Review: Final Report – Measurement of Government Output and Productivity for the National Accounts*. Basingstoke: Palgrave Macmillan.
- Pidd, M. (2003). *Tools for Thinking: Modelling in Management Science*, 2nd edn. Chichester: John Wiley.
- Pollitt, C., and Bouckaert, G. (2004). *Public Management Reform: A Comparative Analysis*, 2nd edn. Oxford: Oxford University Press.
- Public Administration Select Committee (2003). *On Target? Government by Measurement*. London: HMSO.
- Ridgway, V. F. (1956). Dysfunctional consequences of performance measurements. *Administrative Science Quarterly*, 1(2), 240–7.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Smith, P. C. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18(2–3), 277–310.
- Steele, J. (2003). *Involving People in the Public Disclosure of Clinical Data*, report on research with user organisations and patients. London: Nuffield Trust.

PART V

Performance measurement – emerging issues and enduring questions

The final part of the book contains three contributions that both look to the future and explore some of performance measurement's enduring themes. Part V begins with a return to the classic question of whether pay for performance really motivates employees. It then continues by exploring how and why measurement works in some settings, even when it appears that it should not. The final chapter turns to the question of predictive performance measurement – exploring whether it is feasible to develop forward-looking measurement systems.

The first contribution, by Margit Osterloh and Bruno Frey, addresses the question of whether pay for performance really works. Osterloh and Frey review the literature and the results of a variety of field studies that explore whether intrinsic or extrinsic motivation is more powerful. They seek to develop a framework that allows practitioners to ensure that they achieve an appropriate balance of intrinsic and extrinsic methods of motivation. At the heart of their framework is the argument that “extrinsic motivation is sufficient when the work is routinized and performance is easy to measure, while intrinsic motivation is necessary when labor contracts are characterized by a high degree of incompleteness as well as ambiguity”.

The second contribution, from Rob Austin and Jody Hoffer Gittell, builds on Osterloh and Frey's chapter by identifying three basic premises that are taken for granted by most designers of measurement systems: performance should be clearly defined, performance should be accurately measured and reward should be contingent upon performance. Austin and Gittell then present a series of vignettes illustrating how high performance can result even when these basic principles are contravened. This leads them to argue

that there are two different forms of performance measurement. In the first, performance measures are used as part of a management control system that is tightly aligned with an extrinsic reward system. In the second, performance measures result in behavioural modification through ambiguity and intrinsic motivation.

The third contribution, from Thomas Ahrens and Chris Chapman, presents an in-depth case study of a large restaurant chain. Over a four-year period the case company continually sought to improve its measurement systems, yet failed to do so. In spite of this the case company, which was competing in a highly turbulent environment, was able to grow profitability and market share during the same time period. These contrasts lead Ahrens and Chapman to ask why the company was successful when it had a measurement system that was perceived to be ineffective. At the heart of Ahrens and Chapman's thesis is the assertion that the changes that head office sought to make to the measurement system were, in fact, inappropriate. Head office appeared to want a more detailed measurement system that would allow it to take more control over the delivery of service in the local operations. The local operations recognized, however, that they had to tailor the service they were delivering to their clients depending upon the situation and circumstances. Hence, having a tightly coupled and closely prescribed measurement system would have been wrong for the organization given the market in which it was competing.

Together, these three chapters in this final part of the book raise some important and provocative questions that might usefully feed future debate and research on performance measurement.

Does pay for performance really motivate employees?

Margit Osterloh and Bruno Frey

Introduction

Variable pay for performance may undermine employees' efforts, as rewards crowd out intrinsic motivation under identified conditions. A bonus system then makes employees lose interest in the immediate goal. Moreover, monetary incentives in complex and novel tasks tend to produce stereotyped repetition, and measurement is often dysfunctional. Therefore, intrinsic motivation is crucial for these tasks. However, for some work extrinsic incentives are sufficient. We offer a framework for how managers can achieve the right balance between intrinsic and extrinsic motivation.

Variable pay for performance and motivation

Variable pay for performance has become a fashionable proposal over the last few years, in private companies as well as in the public sector. Many firms have given up fixed salaries and have moved to pay their employees in a variable way. Firms try to match payment to objectively evaluated performance. It is reflected in such popular concepts as stock options for managers and various types of bonuses. In the public sector, efforts to raise productivity in the wake of new public management have also resulted in attempts to adjust variably the compensation of public employees for their performance. This means that firms and public administrations increasingly rely on price incentives – i.e. on extrinsic motivations.

We argue in this chapter that variable pay for performance under certain conditions has severe limitations. In situations of incomplete contracts – and these dominate work relationships – an incentive system based only on monetary compensation for work is insufficient to bring forth the performance required. In many situations monetary incentives even reduce performance. Work valued by the employee for its own sake or for fulfilling

personal or social norms is often indispensable. These values or norms may be undermined, or even destroyed, by offering monetary incentives.

Our basic message is that focusing on money as an incentive scheme with complex tasks causes problems. Complex tasks are a typical feature of knowledge-intensive companies, which today comprise the most rapidly growing segment of the economy. In contrast, variable pay for performance (e.g. via piece rates) is adequate only for simple jobs. For complex tasks, monetary rewards are no substitute for good management: relying solely on money is too simple to motivate people in these circumstances. Successful management consists in choosing wisely from among the many different possibilities to evoke interest in the work – i.e. raising intrinsic motivation. This can be achieved by establishing personal relationships within the firm, strengthening participation and securing procedural justice. All serve to communicate to the employees recognition and appreciation of their work.

We first clarify the underlying concepts of extrinsic and intrinsic motivation. On the basis of theoretical and empirical evidence, the following section demonstrates that intrinsic and extrinsic motivation are not additive. Rather, there is a systematic dynamic relationship between the two, called “crowding effects”. In particular, monetary compensation can crowd out the intrinsic motivation to work for one’s own sake. In the following section we discuss when and why intrinsic motivation is needed. However, as argued in the next section, intrinsic motivation sometimes has disadvantages for the organization. Hence an important task of management is to produce the right mixture of motivations. The chapter concludes that the current fad for variable pay for performance is ill-founded.

Intrinsic and extrinsic motivation

Extrinsic motivation exists when employees are able to satisfy their needs indirectly, most importantly, through monetary compensation. Money as such does not provide direct utility but serves to acquire desired goods and services (de Charms, 1968; Deci, 1975; an extensive survey is given by Heckhausen, 1991, chap. 15). Extrinsically motivated coordination in firms is achieved by linking employees’ monetary motives to the goals of the firm. The ideal incentive system is strict “pay for performance”.

In contrast, motivation is *intrinsic* if an activity is undertaken for one’s immediate need satisfaction. Intrinsic motivation “is valued for its own sake

and appears to be self-sustained” (Calder and Staw, 1975, 599; Deci, 1975; Deci and Ryan, 1980). Intrinsic motivation can be directed to the activity’s flow (Csikszentmihalyi, 1975), to a self-defined goal, such as, for example, climbing a mountain (Loewenstein, 1999), or to the obligations of personal and social identities (March, 1999, 377). The ideal incentive system resides in the work content itself, which must be satisfactory and fulfilling for the employees. It follows that, “if you want people motivated to do a good job, give them a good job to do” (Herzberg, as quoted by Kohn, 1993, 49).

Intrinsic motivation is emphasized by the behavioural view of organization. This approach has a long tradition in motivation-based organization theory (Argyris, 1964; Likert, 1961; McGregor, 1960). More recent examples are the critics of transaction cost theory (e.g. Ghoshal and Moran, 1996; Donaldson, 1995; Pfeffer, 1997), as well as the literature on psychological contracts (e.g. Morrison and Robinson, 1997; Rousseau, 1995). They emphasize intrinsic motivation and identification with the firm’s strategic goals, shared purposes and the fulfilment of norms for its own sake.

Intrinsic motivation is dealt with by only a few authors in economics; examples are trust (Arrow, 1974), sentiments (Akerlof and Yellen, 1986; Frank, 1992), firm loyalty (Baker, Jensen and Murphy, 1988), managerial incentives (Güth, 1995) and implicit contracts or norms (Akerlof, 1982). Some economists admit the existence of intrinsic motivation but leave it aside because it is difficult to analyse and control (e.g. Williamson, 1985). Even if the assumption of opportunism is an “extreme caricature” (Milgrom and Roberts, 1992, 42), it is taken to be prudent to consider a worst-case scenario when designing institutional structures (Williamson, 1996; see also Brennan and Buchanan, 1985, for the case of the constitution) – i.e. dealing only with extrinsic motivation. Transaction cost theory goes a step further by assuming that individuals are opportunistic and seek self-interest with guile. Opportunism is a strong form of extrinsic motivation when individuals are not constrained by any rules. The same assumption is made by principal – agent theory, as is made clear in the recent comprehensive surveys by Gibbons (1998) and Prendergast (1999).

To analyse intrinsic and extrinsic motivation in isolation is normally warranted, and corresponds to a useful division of labour between psychology (focusing on intrinsic motivation) and economics (focusing on extrinsic incentives), so long as they do not depend on each other. However, this dependence has been shown to exist, in a large number of careful experiments undertaken by Deci and his group (see Deci, 1971, 1975, Deci and Ryan, 1980, 1985, and Deci and Flaste, 1995). This evidence has largely

been ignored by scholars in organization theory as well as in human resource management (exceptions are Ghoshal and Moran, 1996, and Pfeffer, 1995a, 51, 1995b). The following section considers this relationship.

Motivation crowding out effects

For a long time it has been taken as a matter of course that extrinsic motivation raises performance. It seemed to be a well-established result in both the psychological (see, for example, Eisenberger and Cameron, 1996) and the managerial literature (see, for example, Blinder, 1990, and Lawler, 1990) that positive reinforcement of a particular action increases the future probability of that action. Mechanisms of instrumental and classical conditioning lead to the *(relative) price effect*, which is fundamental for economics (see Becker, 1976, Stigler and Becker, 1977, and Frey, 1992). The opportunity cost of unrewarded behaviour is raised.

However, rewards crowd out intrinsic motivation under particular conditions. The most important conditions are, first, that the task is considered to be interesting (i.e. there must be an intrinsic motivation in the first place) and, second, that the reward is perceived to be controlling by the recipient. This effect has been called the “hidden costs of reward” (Lepper and Greene, 1978) or “the corruption effect of extrinsic motivation” (Deci, 1975). Frey (1997) has introduced it into microeconomics as the “crowding-out theory”. Extensive surveys are given by Lepper and Greene (1978), Pittman and Heller (1987) and Lane (1991). Kohn (1993) and Deci and Flaste (1995) provide popular applications.

The effect may be illustrated with an experience that many parents have with their children. Consider children who are initially enthusiastic about a task. When they are promised a reward for fulfilling the task they lose part of their interest. Parents who try to motivate their children with rewards to do their homework may be successful in the short run. In the long run, however, the children do their homework *only* if they receive a monetary reward. The crowding out effect has set in. In the worst case, the children are prepared to do *any* housework, such as cutting the lawn, only if they are paid.

It is not only for children that such experiences hold; they can be generalized as well. In particular, this principle also applies to variable pay for performance or bonuses. As a consequence, such reward systems usually, but not always, make employees lose interest in the immediate goal (such as serving the customers) and lower their performance. These insights are

corroborated by theoretical arguments as well as by experimental and field studies.

Theoretical background

The crowding out effect is based on *cognitive evaluation theory* (Deci, 1975; Deci and Ryan, 1985) and on *psychological contract theory* (Schein, 1965; Rousseau and McLean Parks, 1993).

According to cognitive evaluation theory, intrinsic motivation is substituted by an external intervention that is perceived as a restriction on acting autonomously. The locus of control shifts from inside to outside the person (Rotter, 1966). The person in question no longer feels responsible but, rather, attributes responsibility to the person undertaking the outside intervention. This shift in the locus of control does not always take place. Each external intervention – e.g. a reward – has two aspects, a controlling and an informing one. The controlling aspect strengthens perceived external control and the feeling of being stressed from the outside. The informing aspect influences one's perceived competence and strengthens the feeling of internal control. Depending on which aspect is more prominent, intrinsic motivation is reduced or raised (see, for instance, the experiments by Enzle and Anderson, 1993). An undermining effect on intrinsic motivation, called crowding out, is complemented by a positive effect on intrinsic motivation, called crowding in. If a task is at one and the same time extrinsically and intrinsically motivated, the more devalued the attribution of a self-determined action is the more strongly the individuals believe themselves to be subject to outside control (Kruglanski, 1975).

The crowding out effect is stronger with material than with symbolic rewards. It is also larger with expected than with unexpected rewards. When the problems at issue are complicated, the negative relationship between reward and performance is stronger than when the problems are simple (Heckhausen, 1991, chap. 15). In all these cases, it is required that the behaviour was initially perceived to be interesting and therefore intrinsically rewarding (see Calder and Staw, 1975).

According to the theory of psychological contracts, each contract includes an extrinsically motivated (transactional) aspect as well as a relational aspect, directed towards a reciprocal appreciation of intrinsic motivation. If the relational part of the contract is breached, the reciprocal good faith is put into question. Empirical evidence shows (Robinson, Kraatz and Rousseau, 1994)

that the parties to the contract then perceive the employment arrangement to be transformed into a purely transactional contract. For example, when a superior acknowledges an employee's extraordinary effort with a symbolic gift (such as a bunch of flowers), the intrinsic motivation of the employee tends to be raised because he or she feels that his or her intrinsic motivation is appreciated. However, if the employee feels that the superior's gesture serves only in an instrumental purpose, his/her intrinsic motivation is impaired. The bunch of flowers is perceived to be controlling; the relationship is interpreted as being transactional.

The reciprocal appreciation of motives also explains why commands normally crowd out intrinsic motivation more than the use of prices. Commands do not take into account the motives of the recipients, while the price system leaves the choice open as to whether one cares about receiving the reward or not.

An important part of psychological contracts involves perceptions of fairness. Fairness includes both outcome as well as procedural justice. *Outcome justice* requires, firstly, that employees evaluate their salary not in absolute terms but relative to their co-workers (see the equity theory by Adams, 1963). Secondly, the contract between employer and employee is seen as a gift relationship based on norms of reciprocity. If the perceived reciprocity is violated, employees reduce their voluntarily offered super-normal performance (Akerlof, 1982; for experimental evidence, see Fehr, Gächter and Kirchsteiger, 1997). *Procedural justice* means that people are prepared to accept substantial differences in wages if the process of their determination is perceived to follow transparent and fair rules (Kim and Mauborgne, 1991, 1998; Tyler, 1990).

Empirical evidence

There have been such a large number of laboratory experiments on crowding out that a more comprehensive view is needed. Fortunately, the experimental evidence has been the subject of several meta-analytical studies. Wiersma (1992) looks at twenty studies covering 1971 to 1990, and Tang and Hall (1995) look at fifty studies from 1972 to 1992. These meta-analyses support the crowding out theory. This view was challenged by Eisenberger and Cameron (1996), who, on the basis of their own meta-analysis covering studies published in the period 1971 to 1991, conclude that the undermining effect is largely "a myth". However, Deci, Koestner and Ryan (1999), as a

result of a very extensive study, are able to show that these conclusions are unwarranted. This most recent meta-analysis includes all the studies considered by Eisenberger and Cameron as well as several studies that have appeared since then. The sixty-eight experiments reported in fifty-nine articles span the period 1971 to 1997, and refer to no fewer than ninety-seven experimental effects. It turns out that tangible rewards, a subset of which is pay for performance, undermine intrinsic motivation for interesting tasks (i.e. tasks for which the experimental subjects show an intrinsic interest) in a highly significant and very reliable way. Thus, there can be no doubt that the crowding out effect exists and is a robust phenomenon of significant magnitude under the conditions identified.

In real-life situations we have to look at the net outcome, composed of the relative price effect as well as the crowding out effect. This holds because the effect of intrinsic motivation cannot always be neatly separated from extrinsic incentives. When someone is fond of communication with customers for fun, for example, it is always possible to find a corresponding external motive such as selling better. Therefore, it is important to consider field studies that take into account the net effect of the relative price and crowding out effects. There exist several such studies.

- (1) A real-life case for the crowding out effect is provided by blood donations, as argued by Titmuss (1970). Paying donors for giving blood undermines the intrinsic motivation to do so. Though it is difficult to isolate the many different influences on blood supply, in countries where most of the blood is supplied gratis, paying for blood is likely to reduce total supply (Upton, 1973).
- (2) A field study refers to the so-called “Nimby” (not in my back yard) syndrome (Frey and Oberholzer-Gee, 1997). In a community located in central Switzerland, in a carefully designed survey, more than a half of the respondents (50.8 per cent) agreed to have a nuclear waste repository built in their commune. When compensation in monetary terms was offered, the level of acceptance dropped to 24.8 per cent.
- (3) An econometric study of 116 managers in medium-sized Dutch firms showed that the number of hours worked in the company decreased with the intensity of personal control effected by the superiors (Barkema, 1995).
- (4) A large-scale study conducted in 3,860 family businesses found that performance pay is ineffective because it violates a psychological contract, directed on higher-order goals such as affiliation and recognition (Buchholtz, Schulze and Dino, 1996).

(5) Austin (1996) shows, on the basis of interviews with eight experts, that performance measurement is highly counterproductive for complex and ambiguous tasks in computer software development.

To summarize, theoretical considerations as well as empirical evidence from laboratory and field research both strongly suggest that external interventions crowd out intrinsic motivation under the conditions specified. In particular, piece rates, bonuses and other forms of variable pay for performance undermine employees' work ethic, especially in complex jobs in which intrinsic motivation is important and rewards are used in a controlling way. The crowding out effect thus provides a possible explanation for the overwhelming empirical evidence that there is no significant connection between pay and performance, except for simple jobs (Gibbons, 1998; Prendergast, 1999). Even for piece rates applicable to simple jobs, the "literature on incentive plans is full of vivid descriptions of the counterproductive behaviors that piece-rate incentive plans produce" (Lawler III, 1990, 58). The same holds for managerial compensation (for a survey, see Barkema and Gomez-Mejia, 1998), a fact that is admitted even by the proponents of principal-agent theory (e.g. Güth, 1995; Jensen and Murphy, 1990).

In the following section we discuss in a more detailed way when and why intrinsic motivation is needed.

Why intrinsic motivation is needed

Intrinsic motivation is required whenever extrinsic rewards in the form of pay for performance lead to undesired consequences.

Firstly, intrinsic motivation is needed for tasks that require creativity; in contrast, extrinsically motivated persons tend to produce stereotyped repetition of what already works (see Schwartz, 1990, and Amabile, 1996, 1998). In addition, experimental research shows that the speed of learning and conceptual understanding are both reduced when people are monitored. The pressure of sanctions leads to lower levels of learning, and the work is performed in a more superficial way than with intrinsically motivated employees (Deci and Flaste, 1995, 47).

Secondly, intrinsic motivation helps to overcome the so-called multiple task problem (Holmström and Milgrom, 1991; Prendergast, 1999). This applies to cases in which contracts cannot completely specify all the relevant aspects of employee behaviour and its desired outcome. Moreover, it is often not clear to the principals which goals are to be set. Financial goals cannot

always be broken down into operational goals for employees. This problem has led to the recent success of the balanced scorecard concept (Kaplan and Norton, 1996). As a result, contracts offering incentives to reach given goals can give rise to dysfunctional behavioural responses. Agents focus only on the rewarded aspects of the job and disregard the unrewarded ones. Neither do they have sufficient incentives to reflect on the adequacy of the goals they should achieve for the overall success of the firm. Multiple task problems are the subject of incomplete contracts, which are characteristic of employment contracts (e.g. Simon, 1951; Williamson, 1975). Empirical evidence suggests that the outcome of incomplete contracts will not normally be evaluated by variable pay for performance but that firms rely considerably on intrinsic motivation (Austin, 1996).

Thirdly, the transfer of tacit knowledge requires intrinsic motivation. Tacit knowledge cannot be expressed in writing or symbols. In contrast, explicit knowledge can be coded, is easily transferable and multipliable and can be stored in books or diskettes. The distinction entails important consequences with respect to the transfer of knowledge and the kind of motivation required. The transfer of tacit knowledge cannot be measured directly. Hence, when several persons contribute their tacit knowledge, joint output is not attributable to a particular person. In the absence of intrinsic motivation, employees would tend to free-ride (Osterloh and Frey, 2000).

Disadvantages of intrinsic motivation

Intrinsic motivation should serve to support a firm's goals. It is not a goal in itself, therefore. Sometimes a specific intrinsic motivation is disadvantageous – for example, surfing the internet for private pleasure during work hours. Consequently, managers must compare the benefits and costs related to motivating employees intrinsically and extrinsically. Though intrinsic motivation is indispensable, it nevertheless does have disadvantages sometimes.

First, changing intrinsic motivation is more difficult, and the outcome more uncertain than relying on extrinsic motivation, or carrots and sticks. This is the reason why economists as well as managers traditionally prefer a reward and command policy (Argyris, 1998).

Second, intrinsic motivation can have an undesirable content. As history shows, some of the most terrible crimes committed have been, at least in part, intrinsically motivated. Envy, vengeance and the desire to dominate are no less intrinsically motivated than altruism, conscientiousness and love.

They all contribute to immediate satisfaction rather than to achieving externally set goals.

Third, extrinsic motivation enables behaviour to become more flexible. The motivation of volunteers in a not-for-profit organization, for instance, depends strongly on how it differs compared to the organizational goal. A profit-oriented firm, in contrast, does not have to be so much concerned about the personal values of its employees as long as it pays them well and the costs of supervision are low.

The art of producing the right motivation

Motivation is not a goal by itself but should serve to support a firm's goals. Enterprises are not interested in producing *some* kind of intrinsic motivation with their employees, say the joy of stamp collecting. Rather, they must be motivated to perform in a *coordinated* and *goal-oriented* way. To this end, managers must compare the benefits and costs related to motivating employees intrinsically and extrinsically. Further research is needed to spell out more precisely the conditions for the right balance between intrinsic and extrinsic motivation. Nevertheless, several systematic determinants can be identified to help to manage the kinds of motivation required within firms.

Extrinsic motivation is sufficient when the work is routinized and the performance is easy to measure. This condition obtains for simple jobs. In such cases, empirical evidence shows that the price effect increases performance. An example is given by Lazear (1996). He finds that, in a large autoglass company, productivity increases of between 20 per cent and 36 per cent of output were reached when the firm switched from paying hourly wages to piece rates.

This is a striking example, showing that extrinsic motivation may result in considerable efficiency gains in a situation in which the people affected by the external intervention have little or no intrinsic motivation.

Intrinsic motivation is a necessary production factor in the firm when labour contracts are characterized by a high degree of incompleteness as well as ambiguity. In contrast to pure market contracts, labour contracts typically include incompleteness to a high degree (Simon, 1951; Hodgson, 1998). In well-defined situations this incompleteness can be outweighed by commands, and the opportunity costs of unwarranted behaviour can be raised. However, if the description of the tasks to fulfil is incomplete *and* ill-defined,

intrinsic motivation – i.e. interest in the task itself – is the only way to avoid shirking.

According to the theories of cognitive evaluation and psychological contracts mentioned above, the following factors favour a higher level of intrinsic motivation.

- (1) *Personal relationships and communication.* A large number of experiments show that communication strongly raises the intrinsic motivation to cooperate (first Dawes, van de Kragt and Orbell, 1988; more recently Frey and Bohnet, 1995). Even if no communication takes place but people can identify each other, cooperation is increased (Bohnet and Frey, 1999a, 1999b). Personal relationships are a precondition for relational psychological contracts (Rousseau, 1995).
- (2) *Participation.* The greater the possibilities to co-determine the more the employees engage themselves in mutually set goals, and adopt them as their own. Participation thus raises self-determination. Mutually agreed goals – in contrast to exogenously imposed goals – strengthen intrinsic motivation, because the employees are informed about their capacity to perform. As experiments show, this effect takes place only when agreement about the goals serves primarily as self-control and self-obligation. In contrast, perceived external control inhibits creativity when people are pursuing goals (see Schwartz, 1990).
- (3) *Interest in the activity.* Employees are more motivated to work when they are aware of the results of their input, when they are responsible for the outcome and when they consider their work to be meaningful (Hackman and Oldham, 1980). Clearly, self-determination is supported.

According to the crowding out theory discussed above, the following factors determine how strongly intrinsic motivation is undermined.

- (1) *Contingency of reward on performance.* The closer the dependence of a reward on the required performance the more strongly intrinsic motivation is crowded out. This holds provided the perceived controlling effect of rewards is stronger than the perceived informing effect. In that situation, employees feel their self-determination to be curtailed. This is an argument in favour of time-based compensation and against strict forms of pay for performance in situations characterized by high intrinsic motivation.
- (2) *Commands.* A command restricts the perceived self-determination of the people affected more strongly than a corresponding reward would. This distinction between the effect of commands and rewards on motivation has been disregarded in traditional economic theory. Recent research on behavioural anomalies reveals that opportunity costs, in our case not

receiving a reward, are systematically less valued than direct costs of the same magnitude, in our case the punishment following from not obeying commands (e.g. Dawes, 1988; Thaler, 1992).

- (3) *Violation of justice*. Agents who feel that they have been paid unjustly reduce their intrinsic motivation. A large number of empirical studies show that people judge the fairness of their pay relative to other individuals. “It is more critical how their pay compares to the pay of others than what they make in absolute dollars and cents” (Lawler III, 1990, 24).

This corresponds to the above-mentioned fairness considerations.

These factors influencing the level of intrinsic motivation obtaining in the firm demonstrate that managing motivation cannot be achieved solely by monetary rewards. The use of money alone causes serious problems for motivating employees.

Concluding remarks

Our main message is: monetary reward is no effective substitute for managing human resources well. Monetary incentives are entangled in an extensive web of psychological contracts, cognitive self-evaluation and fairness considerations. In our contribution we have shown the traps into which management may fall when applying variable pay for performance.

We have systematically disentangled the complicated web of motivations characterizing relationships in firms. In contrast to relationships in markets, contracts in firms are incomplete, with the exception of simple jobs; only in these cases is variable pay for performance applicable and useful. Performance is easy to measure and can be attributed to specific employees. Moreover, such jobs are typically undertaken for monetary reasons only. Today, however, such jobs are increasingly losing importance, as the competitive advantage of firms depends on knowledge and organizational learning. These tasks are, by their very nature, difficult to monitor. Because they are demanding and offer a high degree of self-determination, these activities both enable and necessitate employees to be intrinsically motivated for the firm's goals. In such positions, variable pay for performance as sole motivator crowds out intrinsic motivation.

This does not mean that monetary rewards are unimportant. To avoid the crowding out effect they must take into account various aspects: compensation should be based on an overall evaluation of employees. This may include bonuses or profit-sharing, so long as procedural fairness and perceived justice

relative to co-workers are observed. In any case, to strengthen intrinsic motivation, monetary rewards must be embedded in detailed feedback discussions. This feedback should not just concentrate on outcome criteria but should also express appreciation for the employee's intrinsic motivation. Such feedback supports relational psychological contracts and reciprocal trust. Employees then feel informed about their competence and their contribution to the firm's goals. Despite the current tendency, the use of variable pay as the only tool for handling performance underestimates the complexity of human motivation.

References

- Adams, J. S. (1963). Towards an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67, 422–36.
- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 84, 488–500.
- Akerlof, G. A., and Yellen, J. L. (1986). *Efficiency Wage Models of the Labor Market*. Cambridge: Cambridge University Press.
- Amabile, T. M. (1996). *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.
- (1998). How to kill creativity. *Harvard Business Review*, 76(5), 77–87.
- Argyris, C. (1964). *Integrating the Individual and the Organization*. New York: John Wiley.
- (1998). Empowerment: the emperor's new clothes. *Harvard Business Review*, 76(3), 98–105.
- Arrow, K. J. (1974). *The Limits of Organizations*. New York: Norton.
- Austin, R. D. (1996). *Measuring and Managing Performance in Organizations*. New York: Dorset House.
- Barkema, H. G. (1995). Do job executives work harder when they are monitored? *Kyklos*, 48, 19–42.
- Barkema, H. G., and Gomez-Mejia, L. R. (1998). Managerial compensation and firm performance: a general research framework. *Academy of Management Journal*, 41, 135–45.
- Baker, G. P., Jensen, M. C., and Murphy, K. J. (1988). Compensation and incentives: practice versus theory. *Journal of Finance*, 43, 593–616.
- Becker, G. S. (1976). *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.
- Blinder, A. S. (ed.) (1990). *Paying for Productivity: A Look at the Evidence*. Washington, DC: Brookings.
- Bohnet, I., and Frey, B. S. (1999a). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behaviour and Organization*, 38, 43–57.
- (1999b). Social distance and other-regarding behavior in dictator games: comment. *American Economic Review*, 89(1), 335–9.
- Brennan, G., and Buchanan, J. M. (1985). *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.

- Buchholtz, A. K., Schulze, W. S., and Dino, R. N. (1996). *Breaking the Psychological Contract: A Social Exchange Theory Perspective of Performance-based Managerial Pay*, working paper. Storrs: University of Connecticut, School of Business Administration.
- Calder, B. J., and Staw, B. M. (1975). The self-perception of intrinsic and extrinsic motivation. *Journal of Personality and Social Psychology*, 31, 599–605.
- Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*. San Francisco: Jossey-Bass.
- Dawes, R. M. (1988). *Rational Choice in an Uncertain World*. San Diego: Harcourt, Brace, Jovanovich.
- Dawes, R. M., van de Kragt, A. J. C., and Orbell, J. M. (1988). Not me or thee but we: the importance of group identity in eliciting cooperation in dilemma situations – experimental manipulations. *Acta Psychologica*, 68, 83–97.
- de Charms, R. (1968). *Personal Causation: The Internal Affective Determinants of Behavior*. New York: Academic Press.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105–15.
- (1975). *Intrinsic Motivation*. New York: Plenum Press.
- Deci, E. L., and Flaste, R. (1995). *Why We Do what We Do: The Dynamics of Personal Autonomy*. New York: Putnam.
- Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–68.
- Deci, E. L., and Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. *Advances in Experimental Social Psychology*, 10, 39–80.
- (1985). *Intrinsic Motivation and Self-determination in Human Behavior*. New York: Plenum Press.
- Donaldson, L. (1995). *American Anti-management Theories of Organization: A Critique of Paradigm Proliferation*. Cambridge: Cambridge University Press.
- Eisenberger, R., and Cameron, J. (1996). Detrimental effects of reward: reality or myth? *American Psychologist*, 51, 1153–66.
- Enzle, M. E., and Anderson, S. (1993). Surveillant intentions and intrinsic motivation. *Journal of Psychology and Social Psychology*, 30, 257–66.
- Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: experimental evidence. *Econometrica*, 65, 833–60.
- Frank, R. H. (1992). Melding sociology and economics: James Coleman's foundations of social theory. *Journal of Economic Literature*, 30, 147–70.
- Frey, B. S. (1992). *Economics as a Science of Human Behaviour*. Dordrecht: Kluwer.
- (1997). *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham: Edward Elgar.
- Frey, B. S., and Bohnet, I. (1995). Institutions affect fairness: experimental investigations. *Journal of Institutional and Theoretical Economics*, 151(2), 286–303.
- Frey, B. S., and Oberholzer-Gee, F. (1997). The cost of price incentives: an empirical analysis of motivation crowding-out. *American Economic Review*, 87(4), 746–55.
- Ghoshal, S., and Moran, P. (1996). Bad for practice: a critique of the transaction cost theory. *Academy of Management Review*, 21(1), 13–47.

- Gibbons, R. (1998). Incentives in organizations. *Journal of Economic Perspectives*, 12(4), 115–32.
- Güth, W. (1995). Shirking versus managerial incentives of chief executive officers (CEOs): a note on a possible misunderstanding of principal–agency theory. *Journal of Institutional and Theoretical Economics*, 151(4), 693–8.
- Hackman, R. J., and Oldham, G. R. (1980). *Work Redesign*. Reading, MA: Addison-Wesley.
- Heckhausen, H. (1991). *Motivation and Action*. New York: Springer.
- Hodgson, G. M. (1998). Competence and contract in the theory of the firm. *Journal of Economic Behavior and Organization*, 35, 179–201.
- Holmström, B., and Milgrom, P. (1991). Multitask principal–agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization*, 7(Special issue), 24–52.
- Jensen, M. C., and Murphy, K. J. (1990). Performance pay and top-management incentives. *Journal of Political Economy*, 98, 225–64.
- Kaplan, R. S., and Norton, D. P. (1996). *Translating Strategy into Action: The Balanced Scorecard*. Boston: Harvard Business School Press.
- Kim, W. C., and Mauborgne, R. A. (1991). Implementing global strategies: the role of procedural justice. *Strategic Management Journal*, 12, 125–43.
- (1998). Procedural justice, strategic decision making, and the knowledge economy. *Strategic Management Journal*, 19, 323–38.
- Kohn, A. (1993). *Punished by Reward: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Boston: Houghton Mifflin.
- Kruglanski, A. W. (1975). The endogenous–exogenous partition in attribution theory. *Psychological Review*, 82, 387–406.
- Lane, R. E. (1991). *The Market Experience*. Cambridge: Cambridge University Press.
- Lawler, E. E., III (1990). *Strategic Pay: Aligning Organizational Strategies and Pay Systems*. San Francisco: Jossey-Bass.
- Lazear, E. P. (1996). Performance Pay and Productivity, Working Paper no. 5672. Cambridge, MA: National Bureau of Economic Research.
- Lepper, M. R., and Greene, D. (eds.) (1978). *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. Hillsdale, NY: Erlbaum.
- Likert, R. (1961). *The New Patterns of Management*. New York: McGraw-Hill.
- Loewenstein, G. (1999). Because it is there: the challenge of mountaineering ... for utility theory. *Kyklos*, 52(3), 315–43.
- March, J. M. (1999). *The Pursuit of Organizational Intelligence*. Oxford: Blackwell.
- McGregor, D. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.
- Milgrom, P. R., and Roberts, J. (1992). *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Morrison, E. W., and Robinson, S. L. (1997). When employees feel betrayed: a model of how psychological contract violation develops. *Academy of Management Review*, 22(1), 226–56.
- Osterloh, M., and Frey, B. S. (2000). Motivation, knowledge transfer and organizational forms. *Organization Science*, 11(5), 538–50.
- Pfeffer, J. (1995a). *Competitive Advantage through People*. Boston: Harvard Business School Press.

- (1995b). Incentives in organizations: the importance of social relations. In Williamson, O. E. (ed.), *Organization Theory: From Chester Barnard to the Present and Beyond*, 72–97. Oxford: Oxford University Press.
- (1997). *New Directions for Organization Theory*. Oxford: Oxford University Press.
- Pittman, T. S., and Heller, J. F. (1987). Social motivation. *Annual Review of Psychology*, 38, 461–89.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7–63.
- Robinson, S. L., Kraatz, M. S., and Rousseau, D. M. (1994). Changing obligations and the psychological contract: a longitudinal study. *Academy of Management Journal*, 37(1), 137–52.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(1), 1–28.
- Rousseau, D. M. (1995). *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*. Thousand Oaks, CA: Sage.
- Rousseau, D. M., and McLean Parks, J. (1993). The contracts of individuals and organizations. *Research in Organizational Behavior*, 15, 1–43.
- Schein, E. (1965). *Organizational Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Schwartz, B. (1990). The creation and destruction of value. *American Psychologist*, 45, 7–15.
- Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica*, 19, 293–305.
- Stigler, G. J., and Becker, G. S. (1977). De gustibus non est disputandum. *American Economic Review*, 67(2), 76–90.
- Tang, S. H., and Hall, V. C. (1995). The overjustification effect: a meta-analysis. *Applied Cognitive Psychology*, 9, 365–404.
- Thaler, R. H. (1992). *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. New York: Free Press.
- Titmuss, R. M. (1970). *The Gift Relationship: From Human Blood to Social Policy*. London: Allen and Unwin.
- Tyler, T. R. (1990). *Why People Obey the Law*. New Haven, CT: Yale University Press.
- Upton, W. (1973). Altruism, attribution and intrinsic motivation in the recruitment of blood donors. In *Selected Readings in Donor Motivation and Recruitment*, vol. II, 7–38. Washington, DC: American Red Cross.
- Wiersma, U. J. (1992). The effects of extrinsic rewards on intrinsic motivation: a meta-analysis. *Journal of Occupational and Organizational Psychology*, 65, 101–14.
- Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*. New York: Free Press.
- (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.
- (1996). Economic organization: the case for candor. *Academy of Management Review*, 21, 48–57.

Anomalies of measurement: when it works, but should not

Rob Austin and Jody Hoffer Gittel

Introduction

One ostensible objective of research on performance measurement is to determine the characteristics of measurement systems that lead to high performance. To this end, descriptive theories of performance measurement hypothesize relationships between work contexts, the design of measurement systems, the aptitudes and behaviour of workers, and the outcomes that result from the work. Normative theories specify principles of design for measurement systems intended to produce desirable outcomes.

Usually, normative theories do not specifically assert that high performance results *only* from performance measurement systems constructed in accord with theory. Nonetheless, high performance achieved via means other than performance measurement should surely operate in a way that is *consistent* with the mechanisms inherent in these performance measurement theories. Furthermore, some theories in use, such as the one summarized by the widely accepted aphorism “What you can’t measure, you can’t manage”, seem to imply that a performance measurement system designed according to certain basic principles is a prerequisite for high performance.

This chapter is simple in its layout and objectives. First, we identify a minimal set of principles for designing performance measurement systems that can be extracted from theories in a wide variety of academic disciplines and applied settings. We arrive at three principles that we believe will seem unobjectionable to most performance measurement proponents. Second, we present four vignettes that describe actual events from the airline industry in which the principles of performance measurement seem to be deliberately subverted in order to produce high performance. It is important to emphasize here that these are not merely examples of high performance arrived at via a means other than performance measurement. Rather, in these vignettes basic principles of performance measurement have been specifically subverted in order to improve performance. That there are exceptions to

general theories is not remarkable. We suggest, however, that some of these exceptions challenge existing theories of performance measurement; and that broadening the theories to encompass phenomena such as those described in the vignettes suggests new options for the use of performance measurement to generate high performance.

Basic principles of performance measurement

Research in performance measurement spans fields as diverse as economics, industrial engineering, organizational theory, psychology, public policy and statistics. Methodological approaches vary greatly. It is nearly impossible to survey the research exhaustively across these fields. We focus here on economics and organizational theory, fields that consider performance measurement issues quite directly, and we cite representative research rather than attempting to be exhaustive.

We have made efforts to derive a “lowest common denominator” set of measurement principles. These are principles that, according to many theories, enhance performance. There are research findings that are not consistent with these principles and we do not claim that they are universal. Rather, we use them as theoretical “tripwires” that, when violated, indicate a need to look deeper for explanations.

These principles do have considerable “commonsense” appeal. They are consistent with many practical approaches to performance measurement. Our field research, in software engineering and health care as well as airlines, leads us to believe that there is a great deal of commonality in the notions that underlie practical approaches to performance measurement. We venture that these very basic principles would seem so innocuous to most implementers of real measurement systems that they are often taken for granted.

Principle one: performance should be clearly defined

To state this principle more completely: *performance should be clearly defined, if not in advance, then in terms of criteria that can be agreed after the fact, if not by a third party, then by the worker and manager.*

This principle is an explicit part of many performance measurement systems. For example, the Software Engineering Institute explains the purpose of its software capability evaluation process, which assigns a quantitative

proficiency rating to software development organizations, in the following terms:

[One] objective is to provide a public process which is defined in advance and for which the contractors can prepare. (Humphrey and Sweet, 1987)

Third-party verifiability and definition in advance are seen as desirable characteristics of such systems not only because they are perceived to increase the fairness of a process but also because they are associated with a lack of ambiguity in what constitutes performance. The assumption here is that, if multiple people can observe an outcome and judge it relative to a pre-specified standard of performance, then agreements can be formed about what performance is, and confidence after the fact is increased that high performance has been achieved. Such clear pre-definitions are actionable. Thus, clear definition of performance is associated with the effectiveness of a measurement system in producing desirable outcomes.

The ability to specify performance criteria *before* work is performed is an assumption in much of the economic literature on performance measurement. The mechanisms that underlie much of agency economics (e.g. Ross, 1973; Holmstrom, 1979; Jensen and Meckling, 1976; Holmstrom and Milgrom, 1991) depend on the formation of “contracts”, which are agreements forged between the agent (worker) and principal (owner/manager) on how the benefits of production will be distributed to the worker, contingent on (often measured) outcomes. Within these theories, a clear pre-definition of what constitutes performance allows agency contracts to be enforced.

Institutional economists have explored the implications of the inability to specify performance *ex ante*. Williamson (e.g. 1975) suggests that difficulties in pre-defining performance are a source of incompleteness in contracts that explain in part the use of organizations rather than markets to coordinate production. An economics literature on property rights (e.g. Hart and Moore, 1989) discusses how such difficulties manifest themselves in ownership arrangements, arguing that the structure of ownership might change when there are difficulties in reaching agreement on what constitutes performance.

A significant literature within agency economics addresses the possibility of “subjective contracts” that do not require up-front agreements about performance (e.g. Baker, Jensen and Murphy, 1988; Baker, Gibbons and Murphy, 1994). “Specifying the correct objective measure of employee performance is often impossible . . . The principal knows, in general terms, what he wants the agent to do, but the range of possible actions that the agent

can take, and the range of possible outcomes, is enormous” (Baker, Jensen and Murphy, 1988). Subjective or relational contracts can be used in place of more formal ones “based on outcomes that are observed by only the contracting parties *ex post*, and also on outcomes that are prohibitively costly to specify *ex ante*. A relational contract thus allows the parties to utilize their detailed knowledge of their specific situation and to adapt to new information as it becomes available” (Baker, Gibbons and Murphy, 1997). Even though performance is not third-party verifiable, parties to subjective contracts are dissuaded from renegeing by the possibility of reprisals in future interactions between principal and agent. The expectation of repeated interaction is therefore a requirement for subjective contracting.

As Baker, Gibbons and Murphy (1994) note, engaging in work without a clear *ex ante* definition of performance seems to require *trust*. Also, though performance need not be defined *ex ante* for subjective contracts, the principal and agent must agree on a characterization of performance *ex post*.

The difficulty of defining performance *ex ante* was addressed in organization theory several decades earlier. In reaction to the dominant bureaucratic model (e.g. Weber, 1890 [1979]; Taylor, 1911; Barnard, 1938), which relied heavily on the pre-specification of performance, the neo-Weberians explored its dysfunctional effects. Defining performance *ex ante* reduces conflict in the organization, they argued, but it also increases the rigidity of behaviour (Merton, 1936; Gouldner, 1954). Employees work in accordance with pre-specified performance rather than seeking to do the right thing under the circumstances that arise. Performance is therefore poorer the greater the pre-specification of performance. Gouldner (1954) suggested that organizations often respond to this decline in performance by substituting close supervision in place of pre-defined performance. However, close supervision results in workplace conflict, he argued, motivating a return to *ex ante* performance specifications. The negative cycle is repeated.

Selznick (1949) explored delegation as an alternative to specifying performance *ex ante*, but noted that delegation has its own pitfalls. Given the authority to manage their own performance, units or individuals tend to develop commitment to goals that may not be conducive to meeting overall organizational goals – a problem that motivated later economic agency theories to insist on *ex ante* definitions of performance.

Cohen, March and Olsen (1972) present a “garbage can” model of organizational activity that assumes ambiguity in the pre-work definition of performance. They too suggest an alternative to defining performance *ex ante*: “Although organizations can often be viewed conveniently as

vehicles for solving well-defined problems...they also provide sets of procedures through which participants arrive at an interpretation of what they are doing and what they have done while in the process of doing it” (Cohen, March and Olsen, 1972, 2). Contrary to traditional organizational theory, both the problems to be solved and the desired solutions to them can be defined in the context of doing the work rather than prior to doing the work.

Principle two: performance should be measured accurately

To state this principle more completely: *as work is performed, performance should be measured in a way that conveys the maximum amount of information possible, so that it can be used to determine the degree to which performance has been achieved.*

This principle seems so obvious to most practitioners that it is often merely assumed rather than stated; but it does manifest itself in discussions of what should be measured. For example, Grady and Caswell (1987) suggest that measurement related to the production of software should begin by establishing a wish list of quantities that might be measured, then reduce that set to a handful of quantities that can be measured accurately (and inexpensively). Another reason why practitioners value accuracy in measurement is that they often aim to use measurement information more broadly, as an input to decision making and to promote organizational learning. The importance of accuracy of information in decision-making inputs and for learning is obvious.

A common analogy is drawn between models of engineering control theory and organizational performance measurement. Engineering control models use differential equations to describe and predict the behaviour of physical systems that use measured feedback to adjust themselves. When these models are applied they produce systems that have self-correcting capabilities that also seem very desirable in organizational contexts. One simple example of an engineering control application is the thermostat commonly used for temperature control. The Software Productivity Consortium uses the thermostat analogy as the basis for its performance measurement system.

An example of a closed loop feedback control system is a thermostatically controlled heating system. The thermostat is set to a certain “set point” temperature which is the input goal ... the heat output is continuously compared to the set point goal ... The

thermostat will turn the heater on for a temperature lower than the set point and off otherwise. Uncertainty can exist in the system in establishing the set point according to uncertain temperature requirements. Also, the temperature-measuring device may be inaccurate. However, the system's operation can be improved by ascertaining the temperature requirements and servicing the thermometer.

The closed loop feedback control model represents the software process ... as a "black box" system with interest focused on the input and outputs at the interfaces of the system. Process and product goals are established ... the process is initiated ... measurements are collected ... and compared to measurement goals. The goals correspond to the set point inputs and the measurements correspond to the outputs of the feedback control system ... the difference between the goal and the measurement is the process variance that becomes the driver of process correction.

This analogy suggests that measurement accuracy is an essential element of a performance measurement system – so essential that the measurement device itself requires regular servicing to ensure accurate measurements. Without accurate measurement, performance can exhibit undesirable behaviours, such as accelerating away from the desired standard. This example also reiterates the importance of principle one: lack of agreement about desired performance must be remedied for the system to operate effectively. The thermostat analogy can be detected (and is often quite explicit) behind many performance measurement systems found in practice.

Measurement accuracy clearly plays a role in the construction of verifiable contracts in economic models. This literature has dealt extensively with the consequences of inaccurate signals in the form of statistical imprecision and bias (e.g. Banker and Datar, 1989). Inaccuracy in measurement is typically predicted to diminish performance in these models. There are, however, notable exceptions. Meyer and Vickers (1997) claim that measurement serves the purpose of assessing the capability of the agent, and that therefore inaccurate (i.e. statistically noisy) measures make the agent work harder to convey information about how capable he or she is. Inaccurate measures, therefore, are expected to generate higher performance. Narayanan and Davila (1998) consider trade-offs between the use of measurement information for performance evaluation and its use for belief revision (e.g. for learning about the effectiveness of a process or machine). They demonstrate that conveying information about evaluation can lead to agent manipulation that makes the information less useful for organizational learning.

Organizational theorists have also argued for the importance of the ability to measure performance accurately. Thompson (1967) predicts that organizations will choose between measures of inputs and measures of outputs based on which performance can be measured with greater accuracy. Ouchi (1979) predicts that the accuracy with which measurement can be accomplished is a primary determinant of the mode of control that will be adopted in an organizational setting. Theorists of quality management have emphasized the importance of detailed measurement for the purposes of improving performance.

However, one of the chief theorists of quality management, W. Edwards Deming, warns that measurement is potentially one of the most harmful things an organization can do with respect to improving performance over time (1986). The value of the information contained in a measure must be counterbalanced against the fear of reprisal. Because of that fear, people will systematically subvert the measurement system, causing a decline in the accuracy of the measures and hence undermining their usefulness. Edmondson (1996) demonstrates the negative effects of fears associated with performance measurement, particularly in connection with the sharing of information. This dysfunction suggests the possibility that a better understanding of performance may in fact result from *less* precise measurements, in particular from measurements that lack the diagnostic power to assign outcomes to individual actors. Thus, more ambiguity in the measurements themselves, as well as in the initial definition of performance, may well enhance performance.

Principle three: rewards should be contingent on measured performance

To state this principle more completely: *there should be a clear linkage between desirable performance and rewards such that desirable performance by the worker leads to rewards for the worker; the fact of this linkage, if not the precise relationship, should be known by both worker and manager before work commences.*

Many practitioner systems are consistent with this principle, as it is the basis for the “pay for performance” systems that have become increasingly popular in recent years (Tully, 1993). Clearly, this principle relies on the first two, in that pay for performance requires clear definition and accurate measurement of performance.

The economic agency literature explicitly examines how pay should be made contingent on performance. There are a wide variety of models

addressing a great diversity of issues – for example, whether compensation might be deferred to encourage workers to remain with a given employer (e.g. Lazear and Rosen, 1981), or whether workers might be overpaid (“efficiency wages”) to make the prospect of losing the job less attractive and therefore motivate higher effort (e.g. Shapiro and Stiglitz, 1984). Others have addressed the issue of whether performance pay should be associated with group or individual performance (e.g. Wageman and Baker, 1997). Piece rates have been found to improve performance (e.g. Lazear, 1996), although, as Prendergast (1999) notes, most of these empirical findings have addressed jobs that are quite simple. The economics literature also suggests that firms with clearer linkage between rewards and performance should attract the better workers (Lazear, 1986). In all these models, there is a vital contingency between measured performance and resulting payments to agents. Without the enticement of payments, agents are expected to perform at a nominal level, if at all.

Some economists have noted the potential for dysfunctional behaviours resulting from pay for performance (Baker, Jensen and Murphy, 1988; Baker, 1992). Nevertheless, performance rewards are present even in the models of subjective contracts (Baker, Gibbons and Murphy, 1994). Although there is no third-party recourse, there are punishments that can be inflicted if principal or agent refuse to live up to the terms of the contract, and rewards that can be granted if the parties do honour the contract.

Organizational theorists have also supported making pay contingent on performance. Taylor (1911) introduced pay for performance as part of his proposal to rationalize the workplace through industrial engineering. March and Simon characterized the organizational literature as supporting the view that “the greater the dependence of monetary reward on performance, the more favorable are the consequences perceived as resulting from a decision to increase production” (1958, 82). Rewards work particularly well when performance can be clearly defined *ex ante*, they stipulate, suggesting that the effectiveness of principle three depends upon effective implementation of principle one.

Some organizational theorists, however, have noted the dysfunctional effects of linking rewards and punishments with measurement (Ridgway, 1956; Blau, 1963; Stake, 1971). Consistent with these concerns, there is a substantial literature that suggests the wisdom of relying on intrinsic rather than extrinsic motivation to accomplish some tasks (see, for example, McGregor, 1960, Ouchi, 1981, and Frey, 1993). McGregor argues that “the typical incentive plan is of limited effectiveness as a method of control if the purpose is to motivate

human beings to direct their efforts toward organizational objectives” (McGregor, 1960, 10). People are motivated more effectively through intrinsic than through extrinsic means, he hypothesizes.

Reliance on intrinsic motivation calls for affecting the preferences of workers so that those preferences are more aligned with the objectives of the organization, rather than simply attempting to control worker behaviour (Barnard, 1938). Reliance on intrinsic motivation also calls for the organization to shape some of its own goals to be more consistent with worker preferences, in what McGregor calls the “selective adaptation” to employee preferences. This approach is a departure from traditional organizational and economic theories, which have tended to treat underlying worker preferences as exogenous rather than as having been shaped by participation in the organization.

Proposals to use intrinsic motivation in organizations have raised the question of whether intrinsic motivation can coexist with performance measurement systems based on extrinsic motivators such as rewards. McGregor (1960), Frey (1993) and others suggest that intrinsic motivation may be “crowded out” by extrinsic rewards. McGregor argues that even a well-managed incentive plan “creates attitudes which are the opposite of those desired [including] indifference to the importance of collaboration with other parts of the organization (except for collusive efforts to defeat the incentive plan)” (1960, 9). Prendergast (1999) argues, however, that conclusive empirical evidence of this phenomenon is sparse.

Some have emphasized that the linkage between measurement and rewards often arises regardless of the intentions of the system designers. Even in the absence of explicit rewards attached to measurement, the fear of reprisal from measurements serves in effect as a negative reward (Deming, 1986). Fear of reprisal in turn leads to the hiding of information and efforts to subvert the measurement system.

Summary

Although the principles we list are not universal, we suggest that they do nevertheless retain appeal, especially in practical settings. Where exceptions to the principles have been identified in research, existing theory has tended to focus on specific exceptions one at a time, and to adjust existing models incrementally in order to accommodate exceptions in isolation. Moreover, exceptions have often been characterized as detrimental influences on

performance, such that we cannot do as well as we would be able to if the principles could be held in force. Especially in practice, the prevalent view seems to be that these principles should be aimed for, that conditions conducive to them should be constructed and that they should be adhered to when the conditions can be constructed. Furthermore, it seems a common assessment that the principles are in fact realizable in most situations.

In this contribution, we describe vignettes in which the principles are seemingly violated not only to add to the prevalent sense that existing frameworks need to be further elaborated but also to explore whether the exceptions, considered collectively, might suggest ways of enhancing performance according to a different model – that is, whether the conditions in which the principles can be met, whether naturally occurring or constructed, may be the only or best way to achieve high performance in most organizational settings. Instead, high performance may occur via a model that is not consistent with these principles.

Performance measurement in the airline industry¹

Standard practice for tracking the cause of departure delays in the airline industry is to assign each delay to the party that caused it. Delays that are caused by gate agents (e.g. failure to check in all customers on time) are assigned to customer service. Delays that are caused by the baggage handlers (e.g. failure to load all bags on time) are assigned to baggage. Delays that are caused by flight attendants (e.g. failure to get all passengers seated on time) are assigned to in-flight. Delays that are caused by cabin cleaners (e.g. failure to get the aircraft cleaned on time) are assigned to cabin cleaning. And so on.

The purpose of this system is to assign accountability for delays accurately, in order to evaluate the performance of individual employees and their managers, to motivate better performance and to improve the departure process over time. According to one manager,

It helps a lot just to keep score. People are naturally competitive. They absolutely need to know the score. Once they know, they will do something about it. Every delay comes to my attention and gets a full investigation . . . The last thing most of them want is the spotlight on them.

¹ These observations are taken from a nine-site, four-airline study of the departure process that focused on the effects of formal organizational practices on coordination among front-line employees, and on performance outcomes (see Gittel, 2000a, 2000b).

A customer service supervisor at another airline explained,

The supervisor has to track down the cause of the delay. With a delay, we'll first talk to the customer service agent and the lead and ask what kind of problems were there. Was it the captain, the flight attendant, the control centre, cleaning, catering, a disabled passenger? Any of these would be [someone else's] delay. They should know what time the cleaners got off. If it was a catering delay, it would be charged to catering or to the flight attendants if they counted wrong.

Field managers were evaluated by headquarters along a clearly defined set of performance dimensions, including departure delays. In several companies, headquarters allocated to each field manager a number of acceptable delays, mishandled bags and customer complaints. Any number greater than the quota was unacceptable. The intent of the system was to focus managerial attention, to motivate good performance and to create a basis for evaluating managers over time.

This system has obvious similarities to the engineering control model described earlier. Feedback provided by measurement is perceived as a way in which the organization self-corrects and improves. The system is predicated on the three principles outlined above: pre-specification of what constitutes good performance; the assumption that measurements will be accurate enough to serve as inputs to decision making and improvement efforts; and linkage of performance to rewards (winning the “competition”) and punishments (having the “spotlight” on you). It is important to note that, through these features, the three principles are *designed* into this system.

Evidence of dysfunction in airline performance measurement

There was evidence of dysfunctional behaviours associated with the airline on-time departures system. Some indicators of this dysfunction are listed below.

- A lack of alignment between pre-specified delay codes and actual events (which were more various than anticipated) caused a proliferation of delay types and also mis-coding of delays. “We have delay codes for when the Pope visits, or if there are beetles in the cockpit,” said a field manager, only half in jest, “but sometimes a problem occurs and we have no code for it . . . So we tag it on the last group off the plane.”
- At some stations, employees reported that delays were coded to weather and air traffic control whenever possible, to shift the onus to outside parties.

- There was often a failure to focus on the actual goals of the departure process. According to a supervisor, “If you ask anyone here: ‘What’s the last thing you think of when there’s a problem?’, I bet your bottom dollar it’s the customer. And these are guys who bust their butts every day. But they’re thinking: ‘How do I keep my ass out of the sling?’”
- Execution of the delay tracking system consumed considerable resources that might have been better spent. One customer service employee observed, “Here . . . the ultimate goal is not the customers. It’s the report card. The attitude was, ‘If they are taking a delay at least it’s not mine,’ so you would sort of forget about it. You spend so much time filling out delay forms and fighting over a delay – just think what we could be doing.” A similar sentiment was expressed at another airline. “There is so much internal debate and reports and meetings. This is time that we could be focusing on the passengers.” Rather than focusing on the process itself, managers tended to focus on meeting their numbers to avoid punishment. One manager complained about being “harassed on a daily basis . . . Headquarters has a performance analysis department that is looking at my MAPS [minimal acceptable performance standards] every day, analysing the station’s performance. Failure to meet MAPS is perceived to result in punitive action.”
- Working relationships between groups suffered as well. “There was always a lot of finger-pointing,” according to a ramp manager at one airline. “Barriers between groups – it all comes down to the delay coding system,” said a station manager at another airline.
- Managers were willing to do what was necessary to meet their performance goals, even if it meant doing things that were not in the company’s best interests, in their judgement. “The field manager is judged on the numbers and not on how he got them,” said the manager of human resources. “He could have used a club for all it matters to his rewards.” The employee relations manager concurred: “All that matters is the numbers – how you achieve them is secondary. This is part of the culture of fear.” According to a field manager, “It is scary to delegate, especially here, where there is a very strong culture toward accountability. This is fine, but the penalties that go along with that accountability make people afraid to take risks.”
- Managers transmitted to the front-line workers the pressures they perceived from headquarters. As a result, employees were well aware of their managers’ performance evaluation system, and how it affected them. “Here you only care about delays,” said a customer service agent.

“Otherwise the little report card won’t look good that week. The ultimate goal is not the customers, it’s the report card.”

These dysfunctional behaviours appeared to be linked to the measurement system, which in turn was based on the three principles outlined above. The following vignettes suggest ways that performance in this setting was enhanced by deliberately subverting these principles.

Vignettes: anomalous high performance

Vignette one: “I don’t know how I’m measured” (violation of principle one)

The airline with the industry’s fewest delays, mishandled bags and customer complaints – the best performer in aggregate – is deliberately vague about the basis for managerial evaluation. When asked, field managers are vague about how their own performance is assessed. “I don’t know,” is one typical response, given with a laugh. “I’ll hear about it if I’m not doing a good job.” “It is watched but there is no fear factor,” says another. “Everybody here is a self-motivator.” Some other observations from employees at this high-performance airline follow.

Another field manager at this airline: I know what the relationship [between headquarters and the station] is usually like, because I worked at [another airline] for 20 years. It’s usually an entrenched bureaucracy between the station manager and the headquarters. It’s nothing like that here.

His assistant manager: Each station is like an entrepreneur. We do what we think is right and talk directly to our executive vice presidents and [the CEO]. They are just a phone call away. If they question something we did today, they will call tomorrow.

Field managers at this airline express a comfort with the relationship with headquarters that contrasts dramatically with the resentment expressed by field managers at other airlines. They describe a dialogue with their superiors, and a flow of information that is focused on identifying problems and finding solutions.

Discussion of vignette one

The high performance of this organization seems anomalous in light of principle one, which calls for clear definition of performance. The fact that employees don’t know how they are evaluated suggests that there are no objective performance contracts. However, both the organizational and

economic literatures have acknowledged that dysfunctional behaviour can result from specifying performance *ex ante*. Dysfunctional behaviours, such as those described at other airlines, occur when measures are not true or complete measures of performance – when measures do not fully specify the performance that is desired in part because desired performance is not fully known *ex ante*.

Given the apparent dysfunction induced by the measurement systems of airlines that are not high performers, the ambiguity in the definition of performance at the high-performing airline might be seen as a defence against dysfunction. Evaluation criteria are left ambiguous because the criteria on which a contract could be agreed are not adequate. Being more specific about performance, then, would lead to managers focusing on certain measures to the detriment of overall performance – just as was described above by employees at lower-performing airlines.

It is possible that performance at this organization results from delegating control to the station level instead of pre-specifying performance. As Selznick (1949) and Simon (e.g. 1991) point out, delegation is an option when performance is difficult to pre-specify. This allows the criteria for decision making to be specified at the level of the organizational unit at the time of action rather than in advance. The statement that “each station is like an entrepreneur” seems consistent with this view.

However, this is not a story about a hands-off relationship. The evidence presented here suggests that managerial performance at this airline is evaluated *ex post*, in the context of multiple factors that could not have been specified in advance. One manager points out that performance is watched. Another says, “I’ll hear about it if I’m not doing a good job.” Yet another says that, if headquarters “questions something we did today, they will call tomorrow”. There is a notion of performance being defined by people at headquarters, but, like field managers themselves, they don’t know it until they see it and hear about it in detail. It is only *ex post* that these factors can be analysed, through frank discussion of specific scenarios and specific decisions that were made. Thus, high performance at this airline seems consistent with the use of subjective contracts.

Reliance by this high-performing airline on ambiguous measures appears to require trust, as predicted by Baker, Gibbons and Murphy (1994). It requires trust by field managers that headquarters was motivated more by a desire to coach and develop than to judge and punish. It requires trust by headquarters that field managers were motivated by a desire to do the right thing, whatever that turned out to be in a particular context. The vignette

gives some indication that this trust is indeed present on both sides: “There is no fear factor. Everyone here is a self-motivator.” Perhaps trust is inspired by the ambiguity of the measurement system itself, however. Field managers can trust headquarters because the measurement system removes fear from the relationship. Headquarters can trust field managers because they are “self-motivators”, driven by intrinsic motivation (McGregor, 1960; Ouchi, 1979).

Hence, it seems that the first principle of performance measurement – that performance should be clearly defined – can be detrimental to performance in situations that might be of considerable prevalence, and that its systematic violation can be conducive to high performance.

Vignette two: “team delay” (violation of principle two)

Management at this high-performing airline has also addressed dysfunctional behaviours among front-line employees. These behaviours, known as finger-pointing and “covering your butt”, have been described by employees in several airlines as being common in the industry. This organization’s solution has been to introduce a new delay code called a “team delay”, as an alternative to the more detailed performance measures that were traditionally used. “We’ve had a team delay for a couple of years now,” says a top manager. “We had too many angry disagreements about whose fault it was.” According to another:

The team delay is used to point out problems between two or three different employee groups in working together. We used to do it – if people were still on the jetway at departure time, it was a station delay. If people were on-board at departure time, it was a flight crew delay. But now if you see everybody working as a team, and it’s a team problem, you call it a team delay. It’s been a very positive thing.

“We could have more delay categories,” says the head of operations, “but we just end up chasing our tail.”

The team delay has become the single most-used delay category for station-controllable delays. Within this airline, the team delay is regarded as an innovation that partly explains why this airline has the best on-time performance in the industry.

Discussion of vignette two

Consider what is being described in this vignette: a company institutes a performance measurement system, one that is conventionally used in its

industry. It then intentionally subverts the same performance measurement system by providing a blanket category – the team delay – that can be used to code delays when there is controversy about who caused the delay. Measurement information is apparently lost. Measurement has become less accurate, less revealing of the underlying phenomena. Principle two seems violated. Interestingly, however, performance seems to have improved.

The earlier accounts of the dysfunctional effects of the performance measurement system at relatively low-performing airlines make clear the rationale behind a team delay. The team delay solves many of the problems of dysfunctional performance measurement. It eliminates reasons for devoting resourcefulness to managing numbers rather than actual performance. Delays occur less frequently because people are focused on getting the plane out rather than trying to avoid being “tagged”.

Nevertheless, the team delay clearly results in a loss of measurement information for the organization. When the team delay is used to code a delay, the measurement system does not capture detailed diagnostic information regarding the causes of that delay. Sometimes the team delay is used when a more specific coding would have been possible. Workers realize that the job of getting the aeroplane out in a difficult situation is made easier when there is tacit agreement that any delay would be coded as a team delay. Hence, sometimes the team delay reflects not collective responsibility for a delay but, rather, the perceived difficulty of a situation before any delay occurs. Faced with a difficult situation, early commitment to a team delay makes all the parties more willing to join in problem solving to avoid the delay, even when the delay subsequently occurs anyway. In approving the team delay, the airline is implicitly approving these behaviours. This performance-enhancing solution is, in effect, a company-sanctioned loss of measurement information.

It is not at all clear, however, that there has been a *net* loss of information due to the use of team delays. By allowing the less precise reporting of station delays, the company in reality has *improved* the flow of information and avoided unproductive conflict. People who are not worried about how measurements would reflect on them personally are more willing to discuss the causes of delays. People are, therefore, better able to learn from previous mistakes. Information exchange is voluntary and unimpeded by the efforts to manage appearances that are rampant at other airlines.

All the theoretical frameworks we know of have at least some difficulty explaining why a company would install an information system, then subvert it to improve performance and communication, all the while continuing to

maintain the system. As has been mentioned, there are economic models that suggest that dysfunction might result from accurate measurement. Narayanan and Davila (1998) find that firms will delegate or set up multiple measurement systems when an agent can manipulate measurements that can be used for both evaluation and for belief revision (e.g. learning). This vignette, however, seems not to provide a pure example of either possible solution. As we will discuss later, creating a performance measurement system and then “removing its teeth” may be a gesture intended to communicate a message to workers.

Vignette three: “I’ll take this delay” (violation of principle two)

Tom Dag, a safety coordinator for another airline, has created his own personal solution to the problems created by performance measurement. It is well known at his airline, he explains, that delays are often tagged on the last party still touching the aircraft when the delay occurs. As a result, he explains, people would hold back from helping when a delay seemed imminent, to avoid being tagged. The result, too often, is that an imminent delay becomes an actual delay. This aggravates him.

Nobody wants to be responsible for a delay. It’s a cover-your-butt situation. Personally, I say put the delay on my tab. I’ll do what I can do. It’s my way of jesting with them. They come and help when I do this, because they know I don’t care about a delay. So they don’t have to worry about it being their delay. They come out of the woodwork to help.

Dag is willing to have the delay attributed to him, regardless of the actual cause. The result, however, is to prevent delays that are occurring simply as a result of people trying to avoid being tagged with a delay.

Discussion of vignette three

As in vignette two, Dag’s solution to the problems created by performance measurement results in a loss of information to the organization. The two situations are in many ways similar, and both appear to violate principle two. Vignette three, is, arguably, more difficult to explain, however. Dag’s “creative” approach to delay coding has not been sanctioned by the company. In fact, it is probably a punishable behaviour, were it known at high enough levels in the organization. In addition, he is more likely to be tagged than he would if he behaved as others do – that is, pulling back when a delay seems imminent.

There are several theoretical incongruities in this situation. The company has instituted a measurement system that seems to reduce performance. An employee has taken on significant risks that others are not willing to take to defuse dysfunction and to improve aggregate performance, not to maximize his own well-being. In a game-theoretic frame, it appears that Dag is choosing a dominated or irrational strategy. Although the apparent presence of anomalous cooperation means that economic models lack explanatory power for this vignette, this kind of behaviour is often observed in social situations. Psychologists (e.g. Dawes, 1991) have noted that a variety of social factors cause individuals to choose irrational behaviour, including verbal communication, promising and group identity.

Non-sanctioned behaviours that seek to *neutralize* the effects of a dysfunctional measurement system by reducing the accuracy of its measurements could be quite prevalent in organizations. It is interesting to consider the effects such behaviours might have on attempts to verify empirically the effectiveness of measurement systems.

Vignette four: “I know you’ll do your best” (violation of principle three)²

Ted Derwa, an executive with a major US auto company, was running late trying to catch a flight to Frankfurt. He was sure that he would make it to the plane on time, but he was worried about his luggage. His whirlwind schedule would take him and a group of fellow executives to a different city every day for two weeks; if his bags didn’t get on the plane, they might never catch up with him. He solved this problem in a novel way.

Standing on the airport pavement, he removed a \$20 note from his wallet and handed it to a nearby baggage handler. “Those are my bags,” said Derwa, pointing. “They need to be on the 7:35 for Frankfurt.” Immediately the baggage handler began to protest. He tried to hand back the \$20. There were several reasons why what Derwa was asking of him was impossible, he explained.

Derwa held up a hand, stopping the baggage handler from talking, and said: “Look, to be honest, I don’t really expect my luggage to get on that

² This vignette is not taken from the same field study as the other three. Rather, it is a recounting of a story told within a company that is the former employer of one of this chapter’s authors. Although it surely classifies as organizational “lore” and is part of the “legend” of a widely admired manager, it is represented as fact within this company; parts of the story were corroborated by other “witnesses” within the company.

flight. But I want you to know, you do seem to me like the kind of person who could do it, if it is possible. So I'd like you to keep the twenty and give it your best shot. If it doesn't work out, I'll understand completely." Then Derwa turned and, without looking back, walked into the terminal, leaving the \$20 with the baggage handler and his luggage on the pavement.

A few minutes later Derwa boarded the flight. Within minutes, the plane pulled away from the gate. There was no way Derwa's bags had made it to the plane (he had barely made it himself). As the aircraft reached the taxiway, however, it suddenly lurched to a stop. Looking out through a window, Derwa saw a vehicle speeding towards the plane. In the driver's seat a man was talking urgently into a hand-held radio. Seated beside him, in the passenger seat, was the baggage handler. Piled in the back of the truck were Derwa's bags. The bags were loaded onto the plane. The small group of executives actually broke into applause at the exceptional service.

Discussion of vignette four

This vignette raises numerous questions that are relevant to performance measurement issues.

- The level of performance exhibited by the baggage handler, with respect to the task defined by Derwa, was exceptionally high.
- The "rational" action for the baggage handler in this situation was to pocket the \$20 note and slide the bags onto the luggage belt to be discovered later. There were no imaginable adverse consequences for this act. In choosing instead to take heroic action to get the bags on the plane, the baggage handler probably forwent tips (possibly totalling more than \$20), and maybe even risked reprimand, for departing from procedures.
- There was no conditional linkage between the reward of \$20 and performance of the task. There was no "carrot" and no "stick".

The biggest question here with respect to common frameworks of measurement performance is: "Why did this work at all?" The answer, it seems likely, has to do with the baggage handler's image of himself. Derwa, with his unconditional gift of \$20 and with his brief words, was making an investment in the baggage handler. The fact that it was unconditional made the investment all the more impressive. It communicated a message that said: "Whether or not you succeed this particular time, faced with this particular set of difficulties, I believe you are a person who can succeed and who is worthy of trust and investment." So, honoured by Derwa, the baggage handler wanted to deserve the honour, in his own evaluation of himself.

Given this honoured status, the baggage handler performed to retain it in his own reckoning.

Although the apparent mechanism at work in this vignette will seem at first familiar – many have experienced being motivated by a desire to retain approval – the behaviours herein are anomalous in the context of almost any theoretical framework. It is tempting to access organization theories on intrinsic motivation or commitment. Even within the context of those literatures, though, the events in this vignette remain remarkable, in that the baggage handler's fervour for performance came about so quickly, based on so little interaction between himself and Derwa. Derwa and the baggage handler were strangers, and remained so after their brief interaction. The rapid development of commitment to a "cause" has been demonstrated in other social science contexts. In a set of experiments performance by Dawes, van de Kragt and Orbell (1988), people exhibited seemingly irrational cooperative behaviours, for which they paid personal penalties, when they were told that they were a group that was competing against another group. Fifteen minutes before the experiment none of the participants had ever met, and they had in common only that they had all randomly selected the same coloured chip from a gym bag. Commitment formation is not a part of most theories of performance measurement, however.

Some additional information that might be added to this discussion is this. Derwa was known for this style of management, in which he would place confidence in his employees and act in their interests in a way that was not contingent on outcomes. In this way, he consistently produced strong performance.

Alternative models of performance measurement

How is it that some organizations consistently achieve high performance without relying on the three traditional principles of measurement? In other words, what is the alternative to the traditional model of performance measurement? We hypothesize that what we have referred to as "anomalous high performance" holds within it the promise of systematically achieving high performance. To see how this might operate, we take a closer look at the principles and the vignettes so as to identify what common threads run through them. The vignettes, although some of them are explicable within existing theory, reveal a possible alternative model when viewed collectively.

Taken together, the principles imply a system of performance that might be described as a “closed loop”. Put another way, in accordance with the engineering control analogy to which we have alluded, actual behaviour is compared to desired behaviour (whether defined before or after performance) in a way that is designed to provide at least the opportunity to correct behaviour via rewards and punishments. The control mechanism is systemic and external to agents.

The common thread in the vignettes, however, is that in each the closed loop is intentionally opened and the opportunity for sanctions is forgone. Making aspects of the measurement environment ambiguous for performance measurement purposes opens the loop. The effect is to make some party to the joint effort vulnerable to opportunistic behaviour by the other. By making performance evaluation ambiguous, the high-performance airline’s management makes itself vulnerable to opportunistic behaviour by workers. The team delay operates similarly. Dag’s scheme for reducing delay involves making himself vulnerable to reprisals. And Derwa quite clearly sets himself up for the loss of \$20 without commensurate benefit. Interestingly, high performance seems to follow when one party acts to make her/himself vulnerable to opportunistic behaviour by others who are also involved.

Our hypothesis is that high levels of performance in organizations result when one party to a joint effort, whether an individual or management collectively, acts unilaterally and explicitly to make him-/herself vulnerable to opportunistic behaviour by others. This behaviour, when it works (and we suspect it does not always), aligns the preferences and enhances the intrinsic motivation of the other party.

The act of inserting ambiguity into a performance measurement situation, of opening the performance measurement loop, is a gesture. The gesture is impressive precisely because it exposes the party making the gesture to the potential for opportunism. It contains the implicit message that the gesturing party thinks so highly of the other parties to the effort that it believes the other parties will not take advantage. Placed in this situation, interestingly, people tend to react by also not behaving in a self-interested manner. In fact, the reaction is often to align the preferences and enhance the motivations of all parties.

This phenomenon is familiar in human experience, and is often equated to notions of *heroism*. Acts are made more heroic by the degree to which they exhibit selflessness – that is, by the degree to which they *lack* the type of behaviour that forms the basis for much performance measurement theory. Further, selfless heroism appears to be contagious, in the sense that observing

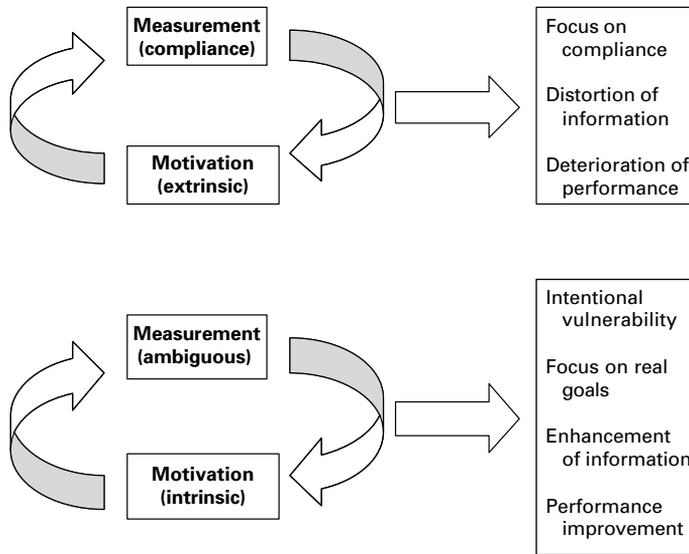


Figure 21.1: Alternative models of performance measurement

a heroic act makes one more likely to behave in similar fashion. Selfless collective action inspired by one party's selfless individual action is a cultural idiom that exhibits itself in inspirational stories and behavioural ideals.

There is a tendency to think of this behaviour as exceptional and not reliable in a systematic way. We hypothesize that this may not be the case. There was little that was obviously exceptional about the baggage handler's situation in vignette four, for example. No one's life was at risk, no grave moral questions were at hand; it was about whether someone's bags got onto an airline flight. Moreover, Derwa was reputed to evoke strong performance consistently in this way. In the case of the baggage handler, he evoked the behaviour almost instantly, despite there being no past or expected relationship between the two parties.

We suggest that there are "meta-models" of performance measurement. One of them, the traditional model that is consistent with the three principles we have identified, is oriented towards compliance (see figure 21.1). Performance is clearly defined, accurately measured, and rewarded based on the assumption that people are extrinsically motivated. This approach to performance measurement tends to evoke extrinsic motivation and fails to evoke intrinsic motivation, further reinforcing belief in the need for a compliance-based approach. The cycle of compliance-based measurement and extrinsically motivated effort tends to give rise to multiple undesired outcomes – namely the deterioration of information quality, the

displacement of employee effort from real organizational goals to those that can be effectively pre-specified and measured, and a deterioration of performance.

The alternative meta-model of performance measurement, which is consistent with some of the literature in economics and organization theory, and which is suggested by our vignettes, is characterized instead by the use of ambiguity to make compliance monitoring difficult. Performance is specified only in a general way, its measurement deliberately fails to capture detailed diagnostic information that would allow precise assignment of responsibility for failure, and rewards are not made conditional on performance. This approach tends to evoke intrinsic motivation and heroic behaviour, further reinforcing the belief that compliance-based measurement is neither necessary nor appropriate.

The cycle of ambiguous measurement and intrinsically motivated effort gives rise to outcomes different from the compliance-based model. First, people tend to focus on the real goals of the organization and on the situation itself, determining in context how these goals might be best achieved. Second, information quality tends to be high, as ambiguous measures call for dialogue across levels regarding what the right thing to do might be in a particular context. Also, because measures fail to pinpoint the responsible party, the information-reducing effects of fear are avoided. Third, the performance of the organization tends to be high relative to organizations using the compliance-based model, and tends to improve over time. There can also be undesirable outcomes from this form of measurement – free-riding may become rampant, for example – but some organizations seem to avoid these.

Consider the following characteristics of an organizational setting:

- the preferences of workers are different from those of the organization;
- the measures are manipulable;
- motivation is intrinsic, at least to some degree; and
- some workers will be inclined to opportunistic behaviour.

These characteristics are, arguably, very prevalent in real organizational settings.

Under these conditions, the compliance-based model will attempt to align the preferences of the workers with those of the organization, but the fact that measures are manipulable will confound these efforts. Intrinsic motivation, if it is not crowded out by extrinsic rewards, will at least be neglected as a means of motivation. Opportunistic tendencies will exhibit themselves in measurement dysfunction. Because the organization seems structured to

reward opportunism, agents will tend to behave opportunistically – that is, they will manipulate measures. They will do so both to protect themselves against unfair evaluation and because organizational incentives seem to encourage this behaviour.

Now consider how the ambiguity-based model operates under the same conditions. By explicitly and intentionally making the organization vulnerable to individual opportunism, the preferences of the individuals who observe this instance of organizational “heroism” will become aligned with those of the organization. Workers will have no reason to manipulate measures because they are not threatening; rather, they will choose instead to volunteer information relevant to the objectives of the organization. Intrinsic motivation will be enhanced. Some workers might free-ride; if they do, there will be performance loss from this source.

We suggest that organizations choose between these meta-models of performance, sometimes without full awareness of the choice. The compliance-based model creates a self-fulfilling prophecy. It leads to the manipulation of measures, which then appear to move in the way expected by managers. Extrinsic rewards lead to extrinsically motivated behaviour and to an expectation that desirable behaviour will be extrinsically rewarded. The implicit assumption that people will not behave “heroically” in the long run is rendered true by the apparent approval of non-heroic behaviour as designed into the performance measurement system. The popularity of pay for performance and other compliance-based techniques may be due in part to this self-fulfilling prophecy. Furthermore, heroic behaviour (e.g. Dag’s “I’ll take this delay”) and the general lack of external validation for the effectiveness of real systems may lead to false conclusions about the effectiveness of compliance-based techniques.

The ambiguity-based model, by contrast, makes use of the contagious nature of selfless behaviour to achieve high levels of performance and to avoid dysfunction. We hypothesize that actions that improve performance via the creation of ambiguity and vulnerability often occur in organizations, overtly in some (e.g. the high-performing airline) but covertly in others (Dag’s airline). Moreover, we suggest that organizations can make systematic use of this model, and that doing so increases the probability that parties within the organization will behave “heroically”. Just as the compliance-based model conveys an expectation that behaviour will be opportunistic, the ambiguity-based model conveys an expectation that it will not be. Worker preferences are endogenous and influenced by the implicit assumptions about tendencies towards opportunism in the design

of performance measurement systems. To make this last point in a more dramatic way: we are what we make of ourselves, and our own performance measurement systems are one powerful way through which we make ourselves.

It seems likely to us that the ambiguity-based model does work within the broad limits of organizational systems intended to protect the organization from the major risks of worker misbehaviour (e.g. embezzlement, theft). These systems tend to involve measurement, but are not performance measurement systems per se. Rather, they are “boundary systems” (Simons, 1996), organizational alarms that detect egregious examples of opportunistic behaviour.

Conclusions

Economic and organization theories have both traditionally supported the three principles of performance measurement put forth in this contribution. More recently, both literatures have suggested weaknesses in these principles, though organization theorists in most cases pointed out these weaknesses earlier than did economists. Both fields have subsequently explored these weaknesses, and in so doing have suggested pieces of an alternative model of performance measurement.

Building upon this previous work and our four vignettes, we have suggested two alternative meta-models of performance measurement. One is based on compliance and evokes extrinsic effort, resulting in goal displacement, distortion of information and declining performance, as well as stronger commitment to the compliance-based approach. The other is based on ambiguity and evokes intrinsic effort, resulting in a focus on real goals, enhanced information and higher performance, as well as stronger commitment to the ambiguity-based approach.

The ambiguity-based model of performance measurement bears resemblance to the high-commitment model of organizations introduced by Walton (1980) and further elaborated by Walton and Hackman (1986). Like our model, the high-commitment model relies on intrinsic rather than extrinsic motivation. As in our model, managers in the high-commitment model elicit intrinsic motivation by voluntarily relinquishing efforts to control. Our model differs, however, in focusing on the role of performance measurement.

We acknowledge that we have not come near to a full description of the ambiguity-based model. In addition, it is impossible to reach conclusive

answers through reasoning from anecdotal information. We maintain, however, that there is something very interesting that asserts itself in these vignettes, and, whatever that is, it deserves more mainstream attention. Moreover, the continued widespread acceptance of the ideal expressed in the three principles of performance measurement seems to us misplaced.

References

- Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3), 598–614.
- Baker, G. P., Jensen, M., and Murphy, K. J. (1988). Compensation and incentives: practice vs. theory. *Journal of Finance*, 43, 593–616.
- Baker, G. P., Gibbons, R., and Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics*, 109, 1125–56.
- (1997). Relational contracts and the theory of the firm. Unpublished manuscript.
- Banker, R. D., and Datar, S. M. (1989). Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, 27(1), 21–39.
- Barnard, C. I. (1938). *The Functions of the Executive*. Cambridge, MA: Harvard University Press.
- Blau, P. M. (1963). *The Dynamics of Bureaucracy: A Study of Interpersonal Relations in Two Government Agencies*, 2nd edn. Chicago: University of Chicago Press.
- Cohen, M., March, J., and Olsen, J. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 17(1), 1–25.
- Dawes, R. M. (1991). Social dilemmas, economic self-interest and evolutionary theory. In Brown, D. R., and Keith Smith, J. E. (eds.), *Recent Research in Psychology: Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs*, 53–92. New York: Springer-Verlag.
- Dawes, R. M., van de Kragt, A. J. C., and Orbell, J. M. (1988). Not me or thee but we: the importance of group identity in dilemma situations – experimental manipulations. *Acta Psychologica*, 68, 83–97.
- Deming, W. E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Press.
- Edmondson, A. (1996). Learning from mistakes is easier said than done: group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science*, 32(1), 5–28.
- Frey, B. (1993). Motivation as a limit to pricing. *Journal of Economic Psychology*, 14, 635–64.
- Gittell, J. H. (2000a). Paradox of coordination and control. *California Management Review*, 42(3), 1–17.
- (2000b). Supporting relational coordination among front-line service workers. *International Journal of Human Resource Management*, 11(3), 517–34.
- Gouldner, A. (1954). *Patterns of Industrial Bureaucracy*. Glencoe, IL: Free Press.
- Grady, R. B., and Caswell, D. L. (1987). *Software Metrics: Establishing a Company-wide Program*. Englewood Cliffs, NJ: Prentice-Hall.
- Hart, O., and Moore, J. (1989). Property rights and the nature of the firm. *Journal of Political Economy*, 98, 119–58.

- Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10, 74–91.
- Holmstrom, B., and Milgrom, P. (1991). Multitask principal–agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24–52.
- Humphrey, W. S., and Sweet, W. (1987). *A Method for Assessing the Software Engineering Capability of Contractors*, Technical Report CMU/SEI-87-TR-.23. Pittsburgh: Software Engineering Institute.
- Jensen, M. C., and Meckling, W. H. (1976). Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 7, 71–9.
- Lazear, E. P. (1986). Salaries and piece rates. *Journal of Business*, 59(3), 405–31.
- (1996). *Performance Pay and Productivity*, Working Paper no. 5672. Cambridge, MA: National Bureau of Economic Research.
- Lazear, E. P., and Rosen, S. (1981). Rank order tournaments as optimal labor contracts. *Journal of Political Economy*, 89(5), 841–64.
- March, J., and Simon, H. A. (1958). *Organizations*. New York: John Wiley.
- Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, 1, 894–904.
- Meyer, M., and Vickers, J. (1997). Performance comparisons and dynamic incentives. *Journal of Economic Management*, 105(3), 547–81.
- McGregor, D. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.
- Narayanan, V. G., and Davila, A. (1998). Using delegation and control systems to mitigate the trade-off between the performance–evaluation and belief–revision uses of accounting signals. *Journal of Accounting and Economics*, 25, 255–82.
- Ouchi, W. G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9), 833–48.
- (1981). *Theory Z: How American Business can Meet the Japanese Challenge*. Reading, MA: Addison-Wesley.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1), 7–63.
- Ridgway, V. F. (1956). Dysfunctional consequences of performance measurements. *Administrative Science Quarterly*, 1(2), 240–7.
- Ross, S. A. (1973). The economic theory of agency: the principal’s problem. *American Economic Review*, 63(2), 4–39.
- Selznick, P. (1949). *TVA and the Grass Roots: A Study in the Sociology of Formal Organization*. Berkeley: University of California Press.
- Shapiro, C., and Stiglitz, J. (1984). Equilibrium unemployment as a discipline device. *American Economic Review*, 74, 433–44.
- Simon, H. A. (1991). Organizations and markets. *Journal of Economic Perspectives*, 5(2), 25–44.
- Simons, R. (1996). *Levers of Control: How Managers Use Innovative Control Systems to Drive Strategic Renewal*. Cambridge, MA: Harvard Business School Press.
- Stake, R. E. (1971). Testing hazards in performance contracting. *Phi Delta Kappan*, 52(10), 583–9.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. New York: Harper and Row.
- Thompson, J. (1967). *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.
- Tully, S. (1993). Your paycheck gets exciting. *Fortune*, 1 November, 83–98.

- Wageman, R., and Baker, G. P. (1997). Incentives and cooperation: the joint effects of task and reward interdependence on group performance. *Journal of Organizational Behavior*, 18(2), 139–58.
- Walton, R. (1980). Establishing and maintaining high commitment work systems. In Kimberly, J. R., and Miles, R. H. (eds.), *The Organizational Life Cycle: Issues in the Creation, Transformation and Decline of Organizations*, 208–91. San Francisco: Jossey-Bass.
- Walton, R., and Hackman, R. (1986). Groups under contrasting management strategies. In Goodman, P. S. (ed.), *Designing Effective Work Groups*, 141–59. San Francisco: Jossey-Bass.
- Weber, M. (1890 [1979]). *Economy and Society*. Berkeley: University of California Press.
- Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

Loosely coupled performance measurement systems

Thomas Ahrens and Chris Chapman

Introduction

This chapter reports on the loosely coupled performance measurement practices of a high-performing restaurant chain (Ahrens and Chapman, 2004, 2007). Over a period of four years we observed a series of initiatives aimed at “tightening up” the performance measurement systems of the case company. However, none of these initiatives resolved the desire for what the finance director towards the end of the research period called “unambiguous performance information”. Head office managers of all grades continued to demand performance measurement systems that allowed more comprehensive and detailed control over the operational decisions of restaurant managers. Whilst it was easy to see in principle how such systems could have been implemented, successive working parties did not change them. This was not for lack of market competition, and it did not result in lower performance. Indeed, managers felt competitive pressures intensify during the research period, and still managed to increase both the market share and the profitability of the case company.

What we were faced with was a high-performing company in a competitive industry that endeavoured to rectify the flaws of its performance measurement system, yet did not. We think that this case holds a lesson for those who are interested in performance measurement system implementation, because it combines high performance with a handling of performance measurement issues that would seem to violate an implicit cornerstone of much of the performance measurement literature. We hypothesize that managers’ “failure” to address the obvious shortcomings of their performance measurement systems may have been related to the management style that enabled the company to perform so remarkably well. Our analysis of this organization suggests the benefits of a certain degree of loose coupling in performance measurement on account of the particular characteristics of its business processes.

Two context dimensions of performance measurement systems

Before analysing these processes in detail, we should explain how loose coupling might benefit performance measurement. By “loose coupling” we mean that specific metrics are only loosely connected with specific managerial responses. This is not the same as bad management. In the case company the link between specific performance metrics and overall restaurant evaluation was ambiguous, but the operational hierarchy was in no doubt about the performance of individual restaurants. Specific sets of metrics informed operational management’s definite views on a particular restaurant. For a different restaurant, however, or at a different point in time, the same measures might well inform a different evaluation and give rise to different responsive actions. This made sense because of the complexities involved in delivering the product. In a full-service restaurant chain standardizing the food is only part of the task. Management also needs to provide an appropriate atmosphere. This is a logistical question, of managing the waiting staff but also of managing the social environment – a task that requires judgement and tact.

Performance measurement in the case company was, therefore, quite different from that in a fast food restaurant chain. In fast food restaurants there is no need to manage waiting staff or the social environment. Restaurant management concentrates on following the centrally optimized rules for assembling standard dishes. Specifications for all operations are detailed and unambiguous (e.g. Leidner, 1993; Love, 1987; Reiter, 1991, chap. 5; Wyckoff and Sasser, 1978, espec. tables 6–3, 6–4). In his study of large fast food chain restaurants in the United States, Bradach (1997) finds that performance evaluation was based on adherence to standards. Head office concerns about violations of brand standards by restaurant managers were so great that profit was not used to evaluate them.

These examples from the restaurant industry suggest that we might usefully distinguish between two dimensions of context in which performance measurement is practised. One dimension is the link between performance metrics and operational responses. This dimension addresses the *use of performance measurement*. After a measurement is taken, what is the appropriate action to follow? Here we distinguish between tightly specified (or programmed) and loosely specified responses. In the former case the purpose of the performance measurement system would be to trigger specific responses by particular readings of performance metrics. For

		Use of performance measurement: looseness of the metric–operational response link	
		Low	High
Local customization of product delivery	Low	1: Programmed standardization (commodity); e.g. fast food restaurant chain	2: Individual standardization (poor administration); e.g. pathological category
	High	3: Programmed customization (mass customization); e.g. financial services	4: Individual customization (tailored product); e.g. full service restaurant

Figure 22.1: Two context dimensions of performance measurement systems

example, a budget overrun for waiting staff would be met with a simple response: stopping overtime.

In contrast, if the performance measurement system is operated in a more loosely coupled fashion, local restaurant managers would be given more freedom to evaluate and justify their budget overrun for waiting staff. They could, for example, explain it as part of their strategy to attract better waiting staff, because they operate in an affluent area where it is difficult to attract local waiting staff for the chain-wide hourly wage. Alternatively, they might explain it as part of a strategy to offer exceptional waiting service to the local clientele, enticing them to purchase high-margin menu items, such as starters and desserts. The key to the successful operation of loosely coupled performance measurement is the autonomy of local restaurant managers, allowing them to vary the parameters of input factors in a way that improves overall financial performance.

The appropriateness of tight or loose coupling depends on who knows best how to handle operations. If the product delivery is standardized, a central office can prescribe operations unambiguously. This is the case in many fast food restaurant chains in which local autonomy is minimal. If, however, product delivery is not wholly standardized, then operations are designed to vary product delivery according to customers' wishes. The second context dimension on which our two examples from the restaurant industry differ is, therefore, the *local customization of the product*. Combining the use dimension with the customization dimension produces the scheme shown in figure 22.1. It contains three viable combinations (boxes 1, 3 and 4) and one that represents failure (box 2).

A fast food restaurant in which customers' only opportunity for customization is the combination of different menu items would occupy box 1.

Attempts at market differentiation by fast food restaurants notwithstanding, a burger is a standard commodity for many consumers. The standardization of operations is comprehensive and allows no looseness of operational responses to the performance metrics of individual outlets.

If there is looseness, the fast food restaurant would find itself in box 2. Standardization, not customization, would still be an objective. Loosely coupled operational responses to defined metrics would, however, introduce uncertainty into the process of selecting the right operational response. In box 2, uncertainty has no productive role to play: uncertainty merely represents a cost to the business process or the consumer. This would constitute a poorly administered bureaucracy, a category, one would hope, that eventually gets eliminated.

Tightly connected metric–operational response links do not preclude product customization if enough information on the customer can be processed. The objective would be to interpret detailed performance metrics, compare them to learned models of customer preferences and select one of the programmed responses. The outcome would be programmed customization, or mass customization. Box 3 therefore comprises organizations that use information on individual customers to tailor offers to them. Recent initiatives in the financial services industry would be examples of programmed customization (Chapman and Gandy, 1997).

Box 4 represents customization without the investment in programmed responses to metrics. The task of customizing lies with individual operators, who tailor the product in response to cues they receive about the customer's preferences. Depending on context, the same set of metrics read on two different occasions can give rise to different responses. Metrics are only loosely coupled to operational responses.

The benefits of loosely coupling metrics and responses

In the case company we observed such loose coupling between performance metrics and operational responses. It provided scope for local restaurant managers and the operational management hierarchy to engage in nuanced discussions of the centrally generated performance information. The potential of loose coupling does not feature prominently in current debates on performance measurement, however. The overall trend in the performance measurement literature is towards advocating systems that integrate measures for various activities according to strategic priorities (Drucker,

1995; Eccles, 1991; MacArthur, 1996; Meyer, 1994; Simons, 1994). Such systems would give rise to detailed analyses of multiple drivers of corporate success, supporting managers in their understanding of operations. The balanced scorecard (Kaplan and Norton, 1992, 1996a, 1996b) is an example. Kaplan and Norton recommend periodic revisions of the performance measures used because no comprehensive system can encapsulate a company's wisdom regarding successful management indefinitely. The coupling between metrics and responses is, however, not discussed.

Neither is the use of performance metrics discussed much in the contingency literature in management accounting. There is awareness that different market conditions, product characteristics and organizational strategies would make different accounting systems designs advisable, but the failure to investigate the ways in which management accounting systems function is one main criticism of this literature (recent reviews include those by Chapman, 1997, Fisher, 1998, and Langfield-Smith, 1997). A combination of the contingency approach with a study of the uses of performance measurement systems would seem promising, because accounting systems model and thereby simplify organizational contexts. By filtering information on the organizational context they buffer decision makers from the uncertainties that arise constantly from the developments surrounding their organization.

Hedberg and Jönsson (1978), in an early paper in this field, speculate about the connections between accounting systems design and use. According to them, in extremely unpredictable environments organizations could benefit from "semi-confusing" information systems. These are systems designed to avoid seducing managers into the illusion that they are keeping tabs on the key variables of the environment when, in fact, those key variables change much more rapidly than internal information systems can be redesigned. Semi-confusing systems present structured views on complex environments without glossing over some of the key contradictions that are present in those environments.

More recently, Ittner and Kogut (1995) explicitly concern themselves with the benefits of keeping the metrics–response link flexible. They discuss the problems that rule-based management processes hold for the development of flexible capabilities. The policies, procedures and measurements that constitute not just formal control systems but also informal norms and evolved rules of thumb constrain the flexibility of managers' responses to contingencies. In companies in which performance measurement readings are taken in diverse contexts, this flexibility can be vital. In our case

company the different localities of the restaurants provided somewhat diverse backgrounds against which managers had to interpret performance measurement information.

The case company

In this chapter we explore these issues on the basis of a longitudinal field study that ran for four years from autumn 1995. The case company is one of the largest chains of wholly owned, full-service restaurants in the United Kingdom. There are no franchises. It has consistently achieved returns on sales of about 20 per cent. During the three years preceding 1998 year-on-year profit growth was above 7 per cent and sales growth was 8 per cent. This growth was attained partly through acquisition but mainly through adding new units. More than 120 restaurants were organized as profit centres, reporting into areas and then regions of operational management. Between 25 and 50 per cent of restaurant managers' remuneration was variable, the variable component depending on controllable restaurant profit, sales and covers growth.

We carried out extensive interviews with all the divisional board of directors, the entire executive committee and selected head office staff. To complement our findings from head office we visited fifteen restaurants, where we interviewed all grades of operational staff and observed operational processes in kitchens and restaurants. Furthermore, we observed meetings between operational area management and restaurant operators, as well as company-wide training sessions on implementing restaurant controls.

The case company brings to the fore issues of loose coupling in performance measurement because of its combination of a key characteristic of chain organizations – standard sets of performance metrics – with the requirement to respond to diverse local operating conditions, as it offered a full waiting service to its customers. Its pricing was such that customers expected a high level of service, adding to the scope for customizing product delivery. We present some data from our field research in the case company in order to demonstrate three points. First, we report on the company's history of management concerns about their performance measurement practices and their failure to make system changes to address those concerns. Second, we describe the operational conditions to suggest that their loosely coupled system was, in fact, advisable. We use one central

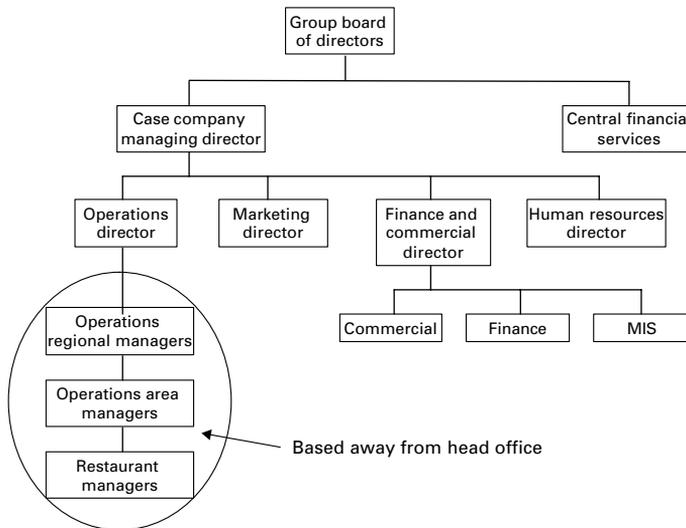


Figure 22.2: Case company organizational structure

performance measure, the food margin, to illustrate this. In doing so we also show how the framework in figure 22.2 can help us to understand managers' discussions of their own organization. Finally, by making links to the main thrust of the performance measurement literature, we suggest some answers to the puzzle concerning managers who kept talking about modifying their performance measurement system when in fact it was serving their purposes well.

A history of "failed" performance measurement system overhauls

Between 1995 and 1999 we tracked ongoing discussions around the case company's performance measurement systems without observing any concrete changes to existing systems. At the beginning of the research period the company had just introduced a new performance measurement initiative, which was given a high profile throughout the organization. It was intended to focus management attention on the "thirteen key tasks". These tasks had been identified as being critical to the success of the case company, and were all matched with a quantified measure and a target level of achievement. This initiative was then dropped after about six months, since it was widely felt that, whilst no one disputed the significance of the issues addressed by the

thirteen key tasks, managers did not think about the management problems facing them in such a fragmented fashion. They found it difficult to think of performance metrics in the abstract, out of the practical context of their specific management problems. None of our interviewees could recall all thirteen key tasks without consulting a written list.

Performance measurement remained on the agenda of the board, however. Their discussions developed a new focus, on the issue of the food margin. The cost of food represented a significant proportion of the overall cost base of the case company, yet it was felt in head office that this central performance measure, which connected the activities of all parts of the organization, was not well understood. Questions were also raised about the measure's calculative implementation through the central management information system. A working party was formed and tasked to resolve this issue.

Various issues were explored by the working party, and they led on to another high-profile initiative within the company. A series of compulsory workshops for restaurant managers was established to reposition the role of centrally generated performance information within the case company. The workshops addressed the concerns of restaurant managers that a reform of performance evaluation systems would turn them into "glove puppets" controlled by head office. The findings of the working party did not result in any changes to the calculative or operational procedures involved in the measurement of the food margin, however.

Some six months after the workshops, at a meeting with the finance director, we were informed that he had set up another working party to clear up once and for all the issues relating to the measurement and management of the food margin. Throughout the whole period of our research within the case company, the reform of performance measurement systems comprised various initiatives that either were aborted completely or produced piecemeal solutions that did little to change the status quo. Towards the end of the research period the finance director denied that the board was pursuing an implicit strategy of keeping the flexibility that a loosely coupled performance measurement system might offer. He reinforced the message that systematic reporting and management was a fundamental aspect of "proper" management, and that the failure to address the issue beforehand was the result of a lack of focus and drive by the board. The finance director was subsequently promoted to another part of the corporate group before the working party could report any findings.

Why nothing changed: the example of food margin reporting

The finance director favoured a technical solution to the difficulties of food margin reporting. In his view the restaurant management loosely coupled operational responses to performance metrics when, in fact, they should have been much more tightly coupled. Since he did not regard the product as highly customized, he wanted to move his organization from box 2 to box 1 (of our figure 22.1). He regarded past management's failure to effect this move as a mistake. Our analysis suggests that box 3 might more appropriately characterize the situation of the organization. Consequently, what he described as bad administration one could reinterpret as beneficial, given the organizational context of high local customization of product delivery. In order to illustrate this line of argument we will explain in more detail the issues around a key "unresolved" concern of the board: the food margin.

The food margin of a dish is the difference between its sales price and the cost of its component raw materials. Individual restaurant budgets were built up from assumptions about targets for covers, sales and the food margin. The case company's management information system allowed a comparison of achieved margin to target margin based on the actual sales mix on a weekly basis. Achieved margin was calculated by dividing food purchases (adjusted for changes in inventory) by actual sales.

Planning for the delivery of food margin targets began anew with the design of every menu (there were up to two new menus per year). Menu design was a complicated process, in which new dishes were designed, tested and costed in the commercial function's test kitchen. The final decision to include or exclude a dish depended on its satisfactory integration into the menu as a whole. A menu had to deliver the desired target food margin percentage while offering a variety of tastes across a range of price points, and conform to the expectations of the brand concept.

Restaurants purchased all food through the central logistics operation at a price fixed for each menu period. The evaluation of the food margin in restaurants was, therefore, based purely on food usage. Price variances were accounted for centrally. Head office prescribed standard procedures for food preparation in the restaurants. The predominant view at head office was that restaurant managers simply had to carry out operations according to the centrally determined parameters and that food margin reporting should police adherence to standards. However, once the food margin reporting system is examined in detail and overlaid with an understanding of restaurant

operation, the interpretation of a food margin deficit or surplus quickly becomes a complicated issue. This tension between expectation and reality was never resolved during our four-year research period.

The standard dish specifications were the subject of mistrust and debate in restaurants. A major concern of restaurant managers was that standards were unrealistic, as they were based on performance in a head office test kitchen that did not make allowances for the difficulty of simultaneously producing many different dishes in a busy and cramped restaurant kitchen. Head office staff felt that restaurant staff paid scant concern to central standards and routinely over-portioned. Restaurant staff, on the other hand, were concerned that standard portions could undermine customer expectations of value by leaving “white space” on the plate. Whereas head office staff were concerned with short-run food margin management, restaurant managers showed more concern for possible damages to long-term profitability.

Our observations in restaurants demonstrated that adherence to food preparation specifications was only a part of food margin management. Margin deficits could arise at any point between unloading food from the delivery lorry to consumption by customers. Restaurant managers felt that deliveries had to be supervised by a relatively senior staff to prevent theft. It was also necessary to check the quality of deliveries, particularly fresh produce such as fruit and vegetables. Theft from junior staff occurred. One restaurant manager estimated that a single disgruntled employee had been responsible for the theft of over £1,000 of inventory over a three-month period.

In the kitchen, chefs prepared some items (such as baked potatoes or chips) in advance in order to reduce guests’ waiting time. Typically, some of these would not be used during a session, but as they could not be kept until the next they therefore contributed to food deficits. In the storage area, failure to implement inventory rotation procedures resulted in food going off and being thrown away. Lapses in concentration during preparation might result in unservable meals. Even appropriately prepared meals could be sent back by customers with particular tastes, or who had simply changed their mind. Staff feeding could amount to more than the specified allowances if left unmonitored. The final source of margin deficits was over-portioning. It could arise from the pace of action in a busy kitchen or the chefs’ reluctance to serve “skimpy” portions. Over-portioning was also perceived to be a deliberate policy of some waitresses, who connected an extra large dessert with a large gratuity.

Adherence to food preparation standards was certainly a significant part of food margin management, but it was intricately tied up with other issues, such as managing guest experiences and expectations and staff expectations. Restaurant managers routinely considered the issue of margin deficits alongside wider issues of restaurant operation, such as staff and customer relations. For example, they clearly felt that theft was a problem. To a certain extent it was seen as an unavoidable aspect of running their kind of operation. When wages were felt to be below a “living wage” a limited amount of theft was even seen as acceptable. In terms of customer relations, one manager consistently ran a food deficit, which he justified in terms of the need to maintain customer expectations coupled with the level of activity in his kitchen, which made a food deficit necessary for the success of his business.

The technical set-up of the food margin calculation implicitly supported restaurant managers who worked the system in these ways. Built into the system were allowances that affected a restaurant’s food margin percentage but that were not connected to the efficient use of raw materials for the dishes. They included allowances for wastage, overestimates of standard food consumption and unmonitored standard allowances for staff feeding. Another possible area in which managers could manage the reporting system was the temporary misreporting of closing inventory.

Analysis and conclusions

As the operations director put it, customers resented being “processed” by the standard chain organization; they demanded to be treated as individuals. The term “restaurant chain” conjures up too simplistic an image of the service process. The standard menu served only as a basis, which restaurant managers had to build on in order to deliver a service tailored to customer expectations. This involved a blending of the production and assembly functions in the kitchen with the service function at the tables. In addition to servicing individual tables, restaurant management also had to keep an eye on the atmosphere in the restaurant as a whole so as to provide an enjoyable eating out experience.

The complexity and interdependence of factors that impacted on restaurant performance in the case company meant that specific metrics could not be tightly coupled to simple prescriptions for management action. Standard sets of metrics read in diverse local operational contexts characterized the performance measurement practice of this full-service restaurant chain. The

operational managers for an area of ten to fourteen restaurants had to adjust their judgement of the same standard set of performance measurements to the local context of every restaurant they visited.

Even looking at a much simpler measure than the food margin, such as sales growth, for instance, we can see the benefits of loose coupling. In the case company, a restaurant's sales growth became the basis of detailed discussions, taking into account the wider context that influenced not just the individual restaurant's levels of activity but also the implications for operational capacity. They would include the latest promotional activities and changes to the central logistics operations, the experience and capabilities of the waiting and kitchen staff, and local competition and demographics. As a result, the specific signal of stagnant sales growth could give rise to a variety of responses, concentrating, for example, on dessert and starter penetration, local marketing, implementing service quality improvements, reducing guest experience time or starting a new training programme for staff.

Our analysis, based on extensive observation of restaurant operation and discussion with restaurant managers, suggested that the company's high level of financial performance was, at least in part, due to the flexibility that the loose coupling within existing performance measurement systems allowed restaurant managers. Whilst reinforcing the necessity of working within the predetermined chain organization format, the performance measurement systems were used by restaurant managers to structure, but not determine, their responses to emerging contingencies and local operational considerations such as staffing level (and capabilities). Restaurant managers were well aware of the drivers of performance within their businesses, but conceived of the various metrics as contributing to performance by setting out a framework within which managers could determine appropriate trade-offs in the light of emerging contingencies.

Restaurant managers' flexibility to engage in trade-offs was constrained by the chain format of the organization. This was borne out in various concerns and comments on the current state of the company and its future. Most managers saw the case company located somewhere between boxes 2 and 4 (figure 22.3). Pessimistic commentators saw the loosely coupled performance measurement system as a shortcoming. Optimists celebrated the restaurant managers' ability to engage in entrepreneurial activity to deliver the challenging targets set by head office. Some restaurant managers recalled how, in the early years, "restaurant managers were king", before head office began to put more emphasis on developing, policing and exploiting the brand concept.

		Looseness of the metric–operational response link	
		Low	High
Product customization	Low	1: Programmed standardization; “Restaurant managers become glove puppets”	2: Individual standardization; “The problem with Food GP is just poor administration”
	High	3: Programmed customization; “We want to move towards intelligent waitresses”	4: Individual customization; “Restaurant managers were king”

Figure 22.3: Different managers categorized the case company differently

Boxes 1 and 3 were referred to in discussion of the possible future of the organization. Box 1 represented the restaurant managers’ dystopian vision of a tight central control regime with no local discretion. Box 3 was various head office managers’ dream of customized restaurant management based on centrally predetermined rules of service. For example, one rule of the brand concept stipulated that, some minutes after serving, waiting staff had to ask customers if they were enjoying their meal. For many customers this enhanced the perception of service, but some customers who were deeply engaged in conversation resented the interruption. One training scheme was designed to teach waiting staff to distinguish when the question was appropriate.

The puzzle remains as to why managers in the case organization kept talking about solving their performance measurement system issues for all those years, apparently without getting closer to a solution. One obvious suggestion would be head office’s inability to make its views felt amongst the restaurants. However, head office management did not have a reputation for weakness in dealing with restaurants. Our analysis would suggest two different points. First, the nature of the product made tight coupling rather difficult to achieve without significant investment in systems of rules and training. In a fast-changing environment such investments might quickly become obsolete.¹ The second point is that, by treating performance

¹ Some recent support for the hypothesis that system “perfection” might be the exception rather than the norm in organizations that need to respond to demand variation in detail comes from Gosselin (1997). In his study of the adoption of activity-based costing he found that formal activity-based costing systems were mainly adopted by bureaucracies. Organizations that inhabited less certain environments and that adopted prospector-type strategies did not implement formal ABC systems but, instead, halted the ABC adoption process at the stage of activity costs analysis.

information as equivocal rather than concrete, the output of performance measurement systems triggered beneficial discussions and explorations of restaurant operation. We observed these in discussions between restaurant managers and head office staff and in training sessions.

The lesson of our case company for those interested in performance measurement systems is that, under conditions that favour the local customization of product delivery, loose coupling in performance measurement may play an important role. For one thing, it can help bring about beneficial discussion of organizational priorities and ways of managing operational inputs to achieve them. It also avoids wasting investment in formal decision rules when fast-moving environments create the risk that they might become obsolete, ossifying the organization's understanding of its situation. Given the pace of change of many business environments, box 3 must be a costly and difficult niche to occupy, and box 1 is not suitable for many organizations. Therefore, box 4 would seem to cover the problems faced by a wide range of organizations. The issue of loose coupling in performance measurement has been largely ignored by the extant literature, and merits much more detailed investigation.

References

- Ahrens, T., and Chapman, C. S. (2004). Accounting for flexibility and efficiency: a field study of management control systems in a restaurant chain. *Contemporary Accounting Research*, 21(2), 271–301.
- (2007). Management accounting as practice. *Accounting, Organizations and Society*, 32(1/2), 1–27.
- Bradach, J.L. (1997). Using the plural form in the management of restaurant chains. *Administrative Science Quarterly*, 41(1), 90–115.
- Chapman, C. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society*, 22(2), 189–205.
- Chapman, C., and Gandy, T. (1997). Datawarehousing: unleashing the power of customer information. *Chartered Banker*, 3(5), 10–42.
- Drucker, P. (1995). The information executives truly need. *Harvard Business Review*, 73(1), 54–62.
- Eccles, R. (1991). The performance measurement manifesto. *Harvard Business Review*, 69(1), 131–9.
- Fisher, J.G. (1998). Contingency theory, management control systems, and firm outcomes: past, present and future directions. *Behavioral Research in Accounting*, 10 (Supplement), 44–64.

- Gosselin, M. (1997). The effect of strategy and organizational structure on the adoption and implementation of activity based costing. *Accounting, Organizations and Society*, 22(2), 105–22.
- Hedberg, B., and Jönsson, S. (1978). Designing semi-confusing information systems for organizations in changing environments. *Accounting, Organizations and Society*, 3(1), 47–64.
- Ittner, C.D., and Kogut, B.M. (1995). How control systems can support organizational flexibility. In Bowman, E.H., and Kogut, B.M. (eds.), *Redesigning the Firm*, 115–80. Oxford: Oxford University Press.
- Kaplan, R.S., and Norton, D.P. (1992). The balanced scorecard: measures that drive performance. *Harvard Business Review*, 70(1), 71–9.
- (1996a). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 74(1), 75–85.
- (1996b). *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.
- Langfield-Smith, K. (1997). Management control systems and strategy: a critical review. *Accounting, Organizations and Society*, 22(2), 207–32.
- Leidner, R. (1993). *Fast Food, Fast Talk: Service Work and the Routinization of Everyday Life*. Berkeley: University of California Press.
- Love, J.F. (1987). *McDonald's: Behind the Arches*. London: Transworld.
- MacArthur, J.B. (1996). Performance measures that count: monitoring variables of strategic importance. *Journal of Cost Management*, 10(3), 39–45.
- Meyer, C. (1994). How the right measures help teams excel. *Harvard Business Review*, 72(3), 95–104.
- Reiter, E. (1991). *Making Food Fast: From the Frying Par into the Fryer*. Montreal: McGill-Queen's University Press.
- Simons, R. (1994). How new top managers use control systems as levers of strategic renewal. *Strategic Management Journal*, 5(3), 169–90.
- Wyckoff, D.D., and Sasser, W.E. (1978). *The Chain Restaurant Industry*. Lexington, MA: Lexington Books.

Index

- Aaker, D.A., 40
- ABC (activity-based costing), 32, 118, 121, 141, 150, 375
- Abela, A.V., 39–52
- Abernathy, W.J., 65, 73
- Abma, T., 415–16, 428
- ABR (activity-based revenue), 8, 118, 121–3
- accountability. *See* responsibility and accountability
- accounting
- financial performance and (*See* financial performance measures)
 - police force performance measurement as form of, 372–3
 - truth in context of, 142
- activity-based costing (ABC), 32, 118, 121, 141, 150, 375
- activity-based revenue (ABR), 8, 118, 121–3
- activity-centric performance measurement, 118, 119, 121–4
- Adams, Chris, xiii, 143
- adaptive planning model, 141–2, 163–4
- beyond budgeting and, 141–2, 174–8
 - complex systems theory, 164–5
 - continuous planning cycles, management through, 167–9
 - coordination of actions across subunits, 173–4
 - front-line teams, planning/strategy/decision making devolved to, 165–7
 - KPI (key performance indicator) reporting, 170–1, 177
 - resource management, 171–3
 - rolling forecast, 169–70, 172
 - “sense and respond” management, shift from “make and sell” paradigm to, 163–4
 - teams, focus of accountability on, 174–8
- Adorno, T.W., 203
- agency theory, 350
- Agle, B., 223
- Ahrens, Thomas, xiii, 428, 432, 477
- Ailawadi, K.L., 53
- airline delay tracking and assignment systems, anomalies in. *See under* anomalies in measurement systems
- ambiguity and uncertainty. *See also* loose coupling in performance measurement systems
- anomalous high performance and, 468–73
 - intrinsic motivation, situations requiring, 440–1, 442
 - in knowledge work measurement, 290
 - in measurement itself, 455
 - in public sector performance measurement, 415–16, 426
 - in specification of performance criteria, 452
- Ambler, Tim, xiii, 39–52, 237, 239
- American Customer Satisfaction Index, 48
- American Express, 172
- American Marketing Association (AMA), 36, 242
- Anderson, B., 65
- Anderson, E.W., 39–52
- Anderson, P., 307
- anomalies in measurement systems, 431, 449.
- See also* loose coupling in performance measurement systems
 - airline industry case studies showing anomalous high performance, 458–68
 - accepting responsibility for delay, leading to less accurate measurement, 465–6
 - delays, tracking and assigning, 458–9
 - perversity and dysfunction in measurement systems, 459–61
 - reward and performance, no clear link between, 466–8
 - “team delay” category leading to less accurate measurement, 463–5
 - vagueness regarding performance criteria, 461–3
- ambiguity model of performance measurement derived from, 468–73
- basic principles of BPM and, 450–8
- accurate measurement of performance, 453–5, 463–5, 465–6

- clear specification of performance criteria, 450–3, 461–3
- compliance-based model of performance
 - measurement derived from, 470
 - reward and performance, clear link between, 455–7, 466–8
- high-commitment model of organizations and, 473
- public sector perversities and dysfunctions, 427–8
- applied mathematics theory of measurement. *See* measurement theory
- Argyris, Chris, 73
- Armstrong, M., 343
- Arora, A., 307
- assets
 - financial performance measurement and, 13, 18
 - marketing performance measurement and, 45–7, 56
- attributability in observation of knowledge work, 287, 288, 289
- Austin, Rob, xiii, 237, 279, 294, 427–8, 431, 440, 449
- axiology, 227
- axiomatic or representational measurement theory, 225–6
- Bacal, R., 342, 343
- Baird, L., 127
- Baker, G.P., 451, 452, 462
- balanced scorecard approach, 147–9
 - financial performance measurement, 23, 27–8, 32, 34
 - institutional/external coherence, lack of, 207
 - key characteristics, 149, 150
 - loose coupling of metric and operational response, 481
 - marketing performance measurement, 37, 48
 - multiple task problem situations, need for
 - intrinsic motivation in, 441
 - operations management, 74
 - shortcomings of, 148
 - strengths and weaknesses, 141
 - widespread adoption of, 143
- Baldrige, Malcolm, and Baldrige Award, 1, 70, 149
- Baldry, C., 354
- Ballantine, J., 150
- Banker, R.B., 283
- Barnard, C.I., 291
- Barwise, P., 47
- Basel Committee on Banking Supervision, 271
- basic principles of BPM, 450–8
 - accurate measurement of performance, 453–5, 463–5, 465–6
 - clear specification of performance criteria, 450–3, 461–3
 - compliance-based model of performance
 - measurement derived from, 470
 - reward and performance, clear link between, 455–7
- BB (beyond budgeting) movement, 30, 34, 141–2, 174–8
- Bebchuk, I., 276
- behavioural psychology. *See* psychology
- benchmarking
 - BPM frameworks working with, 150
 - marketing dashboards, 256
 - operations management, 74
 - performance prism and, 155
 - value of performance measurement and, 218
- Bernardin, H., 354
- Best Factories Award, 75
- beyond budgeting (BB) movement, 30, 34, 141–2, 174–8
- Bicheno, J., 66
- Binmore, K., 296
- bio power, 371, 375
- biotechnology R&D collaboration and innovation performance. *See under* innovation performance measurement
- Bird, Sheila, 409, 410, 411, 412, 413, 423
- Bititci, U.S., 150
- Blau, P.M., 283–5
- Bonoma, T.V., 37
- Bouckaert, G., 408
- boundaries
 - of health care systems, 391
 - of object to be measured, 223
- Bourguignon, C., 127
- Boyle, D., 426
- Boyne, G.A., 409
- BPM. *See* business performance measurement
- Bradach, J.L., 478
- brand equity, 253
- brand valuation, 240, 245, 247
- branding, 46–7, 56
- Brignall, S., 150
- Brignell, J., 426
- British Airways, 171
- Brown, Andrew, xiii, 237, 339
- Brown, M.G., 146
- budgetary control, 29–31, 34
- Buffett, Warren, 177
- Burgman, R., 222
- business or management processes. *See* processes
- business performance measurement (BPM), 1.
 - See also* more specific topics, e.g. marketing performance measurement

- business performance measurement (BPM), (cont.)
- activity-centric, 118, 119, 121–4
 - basic principles of, 450–8
 - causal model of performance, 127–31
 - decision making and performance, 131–2
 - defining, 8, 117–18, 125, 137–8
 - dissatisfaction with current performance measures, 113–14
 - emergent issues and trends, 4
 - firm-centric, 116–21
 - functional perspectives, 4, 7–9
 - future research directions, 431–2
 - indicators vs. measures, 128, 219, 220, 222
 - manageable nature of performance, 136–7
 - as management discipline, 8
 - measurement and performance, relationship between, 134–5
 - measurement, defining, 224–32
(See also measurement theory)
 - organization, performance defined from inside or outside, 132–3
 - performance, defining, 117–18, 125–7
 - practical applications and challenges, 4, 237–8
 - problems of measurement systems per se, 115–16
 - public services, 4
 - reasons for employing, 160, 221 (See also under public sector performance measurement)
 - relativity of concept of performance, 135–6
 - requirements for optimal performance measures, 114–15
 - responsibility issues, 133
 - of subunits, 133
 - timescale of projects, 160
- business process redesign/re-engineering, 151, 154
- Buzzell, R.D., 39
- Cameron, J., 438, 439
- Camp, Robert, 74
- Campanella, J., 70
- Canada
- performance prism framework used in private sector Canadian firm, 159
 - regional health care systems, use of composite indicators in, 397–8
- capabilities
- accuracy of measurement and assessment of, 454
 - in performance prism framework, 155
 - public sector competences, identifying, 410–11, 419, 424
 - TASK (talent, skill and knowledge) differentials in, 238, 280–1, 295–8, 299
- capital, 15, 18–22
- capital asset pricing model (CAPM), 20–1
- Carpenter, M., 308
- cash flow
- branding and, 46
 - as financial performance measurement, 13, 15, 17, 18
 - marketing dashboards, DCF (discounted cash flow) metrics and, 240, 241
 - as marketing performance measurement, 44, 46
 - risk in BPM, 266
- Caswell, D.L., 453
- CATWOE, 418
- causality
- ex post* causal ambiguity in observation of knowledge work, 290
 - performance, causal model of, 127–31
 - results–determinants framework, 146
- cave analogy (Plato), 129
- CbM. See context-based measurement
- Chandler, Alfred, 73
- Chapman, Chris, xiii, 428, 432, 477
- Chartered Institute of Marketing, 36
- Checkland, P.B., 418
- Christensen, C.M., 307
- Church, W. Hamilton, 144
- Chussil, M.J., 39
- citizen focus and police force performance measurement, 369, 376
- Clark, Bruce H., xiii, 8, 36, 37, 39–52
- Clark, Graham, xiii, 238, 318
- CLV (customer lifetime value), 240, 245, 247
- CMO Council, 36
- Coca-Cola, 177
- Cockburn, I., 308
- Codman and Shurtleff, 292, 293
- cognitive evaluation theory, 437
- Cohen, M., 452
- coherence theory, 192–3
- communication and social coherence, 194–6
 - company coherence, 204–5
 - external coherence, 204
 - institutional coherence, 205–8
 - internal coherence, 200–4
 - pragmatic coherence, 198
 - value coherence, 193–4
- collaboration and innovation performance. See under innovation performance measurement
- Collier, Paul, 9, 237, 363, 365
- Collins, J.C., 243, 262, 272
- communication
- of primary teacher professional development needs, 348
 - social coherence and, 194–6
 - theoretical foundations of BPM, 172, 183–4

- companies. *See* organizations
- comparative performance information. *See* cross-organizational measurement
- competences. *See* capabilities
- competition
- balanced scorecard's failure to address, 148
 - market-oriented health care systems, 383
 - results–determinants framework, 146, 148
 - stakeholder approach of performance prism, 151, 155
- complex systems theory, 164–5, 295
- complex tasks and pay for performance schemes, 434
- compliance-based model of performance measurement, 470
- composite indicators. *See* health care systems, use of composite indicators in
- Compstat model of police force performance measurement, 367
- Contandriopoulos, D., 411
- context
- importance of, 319
 - insensitivity to context in observation of knowledge work, 289
 - multiple context dimensions of performance measurement systems, 478–80, 488, 489
- context-based measurement (CbM), 238, 318, 333–4
- association of terms, 328, 329
 - automotive industry case study, 322–3
 - changing coding in, 332–3
 - construction process, 326–30
 - dimensions and terms, determining, 326
 - grammar
 - defined and described, 323
 - designing, 327
 - importance of context for measures, 319
 - language-based approach defined and described, 323–6
 - lexicon
 - classifying and factoring, 324, 327
 - defined and described, 323
 - recording measures, 328–30
 - results, determining, 330–3
 - statements, building and entering, 328, 330
 - structured framework for measuring, use of, 319–21
 - WBS (work breakdown structure), 319–21
- contingency approach, 481
- control issues
- cybernetics, 413–15, 426
 - financial performance of managers (*See under* financial performance measures)
 - health care systems, use of composite indicators in, 390–3
 - in primary school teacher performance, 345–6
 - in public sector performance measurement, 413–16, 419, 421, 426
 - SCM (supply chain management) and control of operations, 90
- Cooper, R., 71
- Cooperative Bank, 2
- Corcoran, F.J., 70
- corporate culture, 91
- corporations. *See* organizations
- correspondence theory, 188–9
- Corvellec, H., 127
- cost centre, operations viewed as, 65–6
- “creative accounting”, 16–17
- creative work, intrinsic motivation required for, 440
- CRM (customer relationship management), 49–50, 87, 102–3
- Crosby, P.B., 67
- cross-organizational measurement
- external coherence of BPM systems, 204
 - in health care systems, 204
 - in operations management, 74–5
 - in SCM (supply chain management), 107–9
- crowding out theory of motivation, 436–7, 457
- empirical evidence of, 438–40
 - theoretical bases for, 437–8
- culture, corporate, 91
- Curtis, B., 296
- customer base
- business process redesign/re-engineering and, 154
 - CRM (customer relationship management), 49–50, 87, 102–3
 - in human activity systems, 418
 - loyalty of, 48–9
 - marketing and (*See under* marketing performance measurement)
 - performance prism and, 152
 - primary school teachers, measuring performance of, 354
 - public accountability of public sector systems, 411–13
 - ROC (return on customer), 240, 241, 251–3, 258
 - satisfaction levels, 47–8
 - SCM and (*See under* supply chain management)
- customer equity, 240, 245, 247, 251, 252
- customer lifetime value (CLV), 240, 245, 247
- customization of product and loose coupling in performance measurement systems, 479, 487–90
- cybernetics, 413–15, 426

- Dag, Tom, 465–6
- Dale, B.G., 70
- Darwinian view of systems, 164
- data envelopment analysis, 39, 67
- Datar, S.M., 283
- Davila, A., 454, 465
- Dawes, R., 226
- Dawes, R.M., 468
- Day, G.S., 39–52
- DCF. *See* discounted cash flow (DCF) metrics
- DeChernatony, L., 47
- Deci, E.L., 435, 436, 438
- decision making
 - in adaptive organizations, 165–7
 - relationship to performance, 131–2
- defining BPM, 8, 117–18, 125, 137–8
- demand management, 88
- Deming, W. Edwards, 67, 168, 409, 455
- Dempster, N., 339
- Denis, J.-L., 411
- Denning, Tom, Lord Justice, 364
- Derwa, Tom, 466–8
- Descarpentries, Jean-Marie, 175
- DHL UK, 156–8
- Diageo, 245
- differential capabilities, 238, 280–1, 295–8
- dimensionalization of measurement, 191, 199–200, 212
- discipline and police force performance
 - measurement, 370–1, 372, 374–5, 376–9
- discounted cash flow (DCF) metrics, 240
 - ROC (return on customer) and, 251–3
 - ROI (return on investment) and, 248
 - suitability of, 240, 241
- dividends, 20
- Drucker, Peter, 73, 74, 342
- DuPont's pyramid of financial ratios, 14–17, 144
- Dutta, S., 309, 313
- dysfunction. *See* entries at perversity and dysfunction
- earned value measurement (EVM), 320
- earnings per share (EPS), 19
- economic agency theory of reward and performance, 455
- economic order quality (EOQ) model, 67, 70
- economic profit models used to account for risk
 - in BPM, 266, 273, 274
- economic value added (EVA[®])
 - financial performance measurement and, 11, 25–6, 28, 32, 33
 - rise in use of, 114
 - risk in BPM, 266, 273
 - SCM (supply chain management) and, 102, 103, 104, 106–7
- education, BPM in. *See* primary school teachers, measuring performance of
- EFQM. *See* European Foundation for Quality Management
- Eisenberger, R., 438, 439
- electronic point of sale (EPOS) systems, 76
- emergent issues and trends, 4
- engineering disciplines, use of measurement in, 319, 453
- Enron, 179
- enterprise resource management (ERM)
 - applications, 319
- enterprise resource planning (ERP) systems, 76
- environmental factors
 - health care systems, use of composite indicators in, 390–3
 - in human activity systems, 419
 - in primary school teacher performance, 345–6
- EOQ (economic order quality) model, 67, 70
- epistemic platform and theoretical foundations of BPM, 189–91
 - data, truth of vs. belief in, 208–12
 - levels of, 201
 - modelling and dimensionalizing, 199–200, 212
- EPOS (electronic point of sale) systems, 76
- EPS (earnings per share), 19
- ERM (enterprise resource management)
 - applications, 319
- ERP (enterprise resource planning) systems, 76
- Eudoxus of Cnidus, 225
- European Foundation for Quality Management (EFQM)
 - award, 1, 70
 - Business Excellence Model, 149, 150
 - financial management performance measurement scheme, 27, 32
- Eusler, Ken, xiv, 8, 125
- EVA[®]. *See* economic value added
- evaluability in observation of knowledge work, 288, 289
- EVM (earned value measurement), 320
- evolutionary view of systems, 164
- ex post* causal ambiguity in observation of knowledge work, 290
- external financial reporting, 18–22
- “face” and risk appetite, 264
- fact. *See* *under* theoretical foundations of BPM
- factor analysis (FA), 388
- Fahey, L., 39–52
- fairness, perception of, 438, 444
- Feder, R.A., 39

- feedback loops
 closed-loop feedback systems, 453
 cybernetics, 413–15
 marketing performance measurement, 54
 PDCA (plan–do–check–act) cycle, 168, 326, 409
 360-degree feedback, 353
- Fiengenbaum, A.V., 67
- Fielding, Michael, 354
- financial performance measures, 7, 11–35
 advantages and disadvantages, 31–3
 assets, 13, 18
 budgetary control, 29–31, 34
 capital, 15, 18–22
 cash flow, 13, 15, 17, 18
 connections between approaches to, 28–31
 “creative accounting”, 16–17
 definition of accounting, 32
 external financial reporting and shareholder
 return on investment, 18–22
 loose coupling of metric and operational
 response, 481
 as mainstay of quantitative approaches to
 BPM, 11
 as major objective of business, 18–22
 managerial motivation and control via, 22–8
 accounting-based techniques, 22, 23–7
 balanced scorecard approach, 23, 27–8, 32,
 34
 budgetary control, 29
 EVA[®], 11, 25–6, 28, 32, 33
 investment centres, 23
 non-accounting performance drivers, 22,
 27–8
 residual income, 25
 responsibility centres, 23
 revenue centres, 23
 transfer price, 24
 non-financial measures vs., 11, 32
 profitability, 13, 15, 17, 18, 19
 ratios, pyramid of, 14–17, 144
 relative vs. fixed targets, 31
 research challenges in, 33–5
 ROI, 12, 18–22
 strategy mapping, 34
 as tool of financial management, 12–18
- firm-centric performance measurement, problem
 of, 116–21
- firms. *See* organizations
- Fisher, I., 210–11
- Fitzgerald, L., 146, 148, 150
- Flaste, R., 436
- Fletcher, A., 222
- Fletcher, C., 354
- Ford, Henry, 66
- forecasting process, 169–70, 172
- Fornell, C., 40
- Foucault, Michel, 363, 369–72, 373, 376
- Fox, A., 350
- frameworks and methodologies, 4, 141–2.
See also specific types, e.g. balanced
 scorecard approach
- CbM, 319–21
- innovation performance measurement,
 306–8
- integrated learning theory of (*See* integrated
 learning theory of truth)
- key characteristics, 149–51
- managerial buy-in, 160
- measurement theory of (*See* measurement
 theory)
- non-core elements, 150
- operations management’s need for, 72–4
- reasons for adopting, 160, 221
- review of, 143–4, 144–9
- SCM (supply chain management)
 GSCF conceptual framework, 84–5
 metrics, framework for developing, 100–9
 timescale of projects, 160
- Fraser, Robin, 141
- Freeman, R.E., 151
- Frey, Bruno, xiv, 431, 433, 457
- Fried, J., 276
- functional perspectives on BPM, 4, 7–9
- funding
 of health care systems, 391
 police force performance measurement and,
 375
 primary school teacher professional
 development and performance
 measurement, 349
- future discounted cash flow (DCF) metrics, 240
 ROC (return on customer) and, 251–3
 ROI (return on investment) and, 248
 suitability of, 240, 241
- Gambardella, A., 307
- gaming, 423, 424
- Garvin, D., 221
- GE Measurement Project, 144
- GEC/AEI merger, 179
- General Electric Company, 27, 271, 272
- Geroski, P.A., 272
- Gerwin, D., 72
- Gibbons, R., 435, 451, 452, 462
- Gillette, 177
- Gittell, Jodi Hoffer, xiv, 294, 427–8, 431, 449
- Gleeson, D., 340
- Global Manufacturing Research Group, 75

- Global Supply Chain Forum (GSCF), 82, 84–5, 109
- goal-setting theory and measurement of teacher performance. *See under* primary school teachers, measuring performance of
- Goddard, Maria, xiv, 237, 383
- Goldstein, H., 347
- Gooner, R., 39
- Gouldner, A., 452
- governmentality and police force performance measurement, 371–2, 375–6, 376–9
- Grady, R.B., 453
- Gray, D., 222
- Greene, D., 436
- grounded theory, 189, 190, 191, 192, 199–200, 212
- Groupe Bull, 175, 176
- Gruca, T.S., 40
- GSCF (Global Supply Chain Forum), 82, 84–5, 109
- Habermas, J., 194
- Hackman, R., 473
- Hall, R.W., 67
- Hall, V.C., 438
- Hammer, Michael, 154
- Handelsbanken, 167, 171, 173, 176–7, 178
- Harris, F., 47
- Hayes, R.H., 65, 73
- health care service in UK. *See under* United Kingdom
- health care systems, use of composite indicators
in, 237, 383–4, 404
advantages and disadvantages, 384–5
boundaries of system, 391
collinearity of variables, 388
combining variables, 389–90
decision rules, using, 390
definition of composite indicator, 383
efficiency of resource management, 393
environmental or uncontrollable factors, adjusting for, 390–3
funding of system, 391
international case studies, 394–403
Canadian regional health care, 397–8
UK NHS (National Health Service), 398–9, 401–2
US Medicare and Medicaid, 394–7, 402–3
WHO (World Health Organization), 399, 401
multidimensionality, need to deal with, 383
organizational objectives, choosing, 386
policy implications, 404
preferences, eliciting, 389–90
process vs. outcome measures, 387
risk adjustments, 392
sensitivity analysis, 393
specific indicators or variables, choosing, 386–8
steps for developing, 385–6
transformation or standardization of chosen variables on commons scale, 388–9
units to be assessed, choosing, 386
value measurement, 389–90
weighting variables, 389–90
- heat charts, 270
- Hedberg, B., 481
- Heller, J.F., 436
- Helmholtz, H., 225
- Henard, D.H., 48
- Henderson, R., 308
- “heroic” model of performance measurement, 468–73
- hierarchy of effects, 43
- high-commitment model of organizations, 473
- high performance with poor performance measurement systems. *See* anomalies in measurement systems; loose coupling in performance measurement systems
- highly skilled workers
creative work, intrinsic motivation required for, 440
knowledge workers (*See* knowledge work measurement)
in public sector performance measurement, 415–16
teachers (*See* primary school teachers, measuring performance of)
- Hofstede, G., 238, 413–15, 416, 426
- Holmström, B., 281, 285, 290, 291, 293
- Hood, C., 413
- Hope, Jeremy, xiv, 141–2, 163
- Hoyer, W.D., 39
- Huff, F., 426
- human activity systems, 418–19, 420
- Husbands, C., 340
- ICs (interactive control systems), 292–3
- IMA (Institute of Management Accountants), 113
- Imai, M., 67
- incomplete contracts
intrinsic motivation and pay for performance, 433, 441, 442
in knowledge work (*See* knowledge work measurement)
- indicators
composite (*See* health care systems, use of composite indicators in)
leading, 136
measures vs., 128, 219, 220, 222
- individual reward incentives and risk, 275–7

- industrially radical innovation, 304
 information asymmetry in observation of
 knowledge work, 285–90
 information availability in using pupil progress as
 measure of teacher performance, 347–8
 information flow structure, 91
 information systems, “semi-confusing”, 481
 informational measurement, 297–8
 innovation performance measurement, 238, 304–5
 biotechnology R&D collaboration sample study,
 310–16
 analysis of patent-based measures, 313–15
 hypotheses of, 310–11
 Poisson regression analysis, use of, 312, 313
 results of, 312, 313
 testing methodology, 311–12
 categories of radicality, 304–5, 315
 frameworks and methodologies, 306–8
 importance of studying, 305–6
 industrially radical innovation, 304
 managerial implications of, 316
 organizationally radical innovation, 304
 patent citation analysis, 238, 304, 315
 advantages of, 307
 in biotechnology sample study, 313–15
 citation weighting, 308–9
 limitations of, 308
 technologically radical innovation, 305
 user-radical innovation, 304
 inseparability in observation of knowledge work,
 289
 Institute of Management Accountants (IMA), 113
 institutional coherence of BPM systems, 205–8
 integrated learning theory of truth, 196–7
 company coherence, 204–5
 data, belief in vs. truth of, 208–12
 designing valid system from perspective of, 198
 epistemic platform, modelling and
 dimensionalizing, 199–200, 201, 212
 external coherence, 204
 grounding in reality and life world of
 performance unit, 198–9
 institutional coherence, 205–8
 internal coherence, 200–4
 proactive truth, real truth, and learning,
 212–14
 purpose and objectives of BPM systems,
 pragmatic coherence to, 198
 integration, 184, 196–7
 interactive control systems (ICs), 292–3
 International Manufacturing Strategy Survey, 75
 intrinsic motivation. *See under* pay for
 performance and motivation
 inventory turns, 92–5
 investment centres, 23
 iQuanta system and police force performance
 measurement, 367, 374, 380
 Ittner, C., 481
 Jacobs, Rowena, xiv, 237, 383
 Jacobson, R., 40
 Japan’s emergence as economic power, 67
 Jatur, S., 222
 Jencks, S.F., 394
 Jensen, M.C., 76, 275, 451
 JIT (just-in-time), 65, 71
 Johnson and Johnson, 243, 292, 293
 Johnson, H.T., 71, 129
 Jönsson, S., 481
 Juran, Joseph M., 67
 just-in-time (JIT), 65, 71
 kaizen, 67, 69
 Kaplan, R.S., 71, 73, 350. *See also* balanced
 scorecard approach
 Katila, Riitta, xiv, 238, 304
 Kay, J., 272
 Keegan, D.P., 145, 149
 Kelleher, Herb, 174
 Kelvin, Lord, 134, 203
 Kennerly, Mike, xiv, 141, 143
 Kettl, D.F., 339
 key performance indicators (KPIs), 170–1, 177
 key result indicators (KRIs), 170
 King, Lord, 171
 Knemeyer, Michael, xiv, 8, 82
 knowledge and power, police force performance
 measurement as, 363, 369–72, 379–80
 knowledge work measurement, 237, 279–81
 attributability, 287, 288, 289
 context insensitivity, 289
 evaluability, 288, 289
 ex post causal ambiguity, 290
 future research challenges, 298–9
 informational measurement, 297–8
 inseparability, 289
 measurability, 287, 288, 289
 motivation, 280, 290–4
 motivational measurement, 297–8
 observability, 279–81, 281–90 (*See also*
 observability and BPM)
 pay for performance, 434
 practical implications of, 300
 reliability, 289
 TASK (talent, skill and knowledge) differentials,
 238, 280–1, 295–8, 299
 use of term, 279
 Koestner, R., 438

- Kogut, B., 481
- Kohn, A., 436
- Kortum, S., 308
- KPIs (key performance indicators), 170–1, 177
- Krafft, M., 39
- Krantz, D., 225
- KRIs (key result indicators), 170
- Kudlow, Larry, 251
- Lambert, Douglas, xiv, 8, 82
- Land Securities plc, 273
- Lane, P., 316
- Lane, R.E., 436
- Langley, A., 411
- language-based approach of CbM, 323–6.
See also context-based measurement
- Larkey, Pat, xiv, 237, 279
- Latham, Gary, 76, 342, 344, 345, 347, 357
- law enforcement. *See* police force performance measurement
- Lawler, E.E., III, 440, 444
- laws and regulations
 risk management, 262
 stakeholders, legal and regulatory communities as, 152
- Lazear, E.P., 442
- leading indicators, 136
- league tables, 424–5
- learning approach to validity and truth, 196–7.
See also integrated learning theory of truth
- learning environments, BPM in. *See* primary school teachers, measuring performance of
- learning, performance measurement as supporting, 421
- Leasco–Pergammon affair, 179
- Leavitt, B., 291
- Lebas, Michel, xiv, 8, 125
- lender ratings used to calculate risk in BPM, 268–9
- Lepper, M.R., 436
- Lerner, J., 307
- Likert scale, 226
- Likierman, Andrew, xv, 237, 261
- Linstone, H., 224
- local customization of product and loose coupling in performance measurement systems, 479, 487–90
- Locke, Edwin, 76, 344, 345, 357
- logic and theoretical foundations of BPM, 182, 192–3
- loose coupling in performance measurement systems, 432, 477
 defined, 478
 local customization of product, 479, 487–90
 multiple context dimensions of measurement systems and, 478–80, 488, 489
- operational responses and performance metrics, 478–9
- in research literature, 480–2
- restaurant chain case study, 482–3
 differing managerial views of company, 488, 489
 failed performance measurement reforms in, 483–4, 489
 food margin reporting system, 485–7
 organizational structure, 483
- loyalty of customers, 48–9
- Lubatkin, M., 316
- MacNeal, K., 179
- Macro Process Model, 146, 147
- Magretta, Joan, 277
- “make and sell” paradigm, shift to “sense and respond” management from, 163–4
- Malcolm Baldrige National Quality Award, 1, 70, 149
- management
 adaptive planning (*See* adaptive planning model)
 BPM as disciple of, 8
 financial performance as means of motivating and controlling (*See under* financial performance measures)
 innovation performance measurement, implications of, 316
 performance amenable to, 136–7
 police force shift from measurement to, 376
 support for BPM projects, 160
- management by objectives, 342
- management or business processes. *See* processes
- manufacturing flow management, 88
- Manufacturing Future Survey, 75
- manufacturing, generally. *See* operations management
- March, J.G., 291, 452, 456
- market share, 44
- market value, concept of, 210–11
- marketing accountability, 242
- marketing dashboards, 237, 239–41, 257
 changing metrics, 245
 DCF metrics (*See* discounted cash flow (DCF) metrics)
 definition of marketing for purposes of, 239
 future research in, 256–7
 how many and which metrics to include, 243–5, 253–6
 importance of performance evaluation in marketing, 241–2
 multidimensionality, importance of, 240, 242–4

- planning and performance evaluation metrics
 - differentiated, 253–6, 257
 - practical application of, 256–7
 - ROC (return on customer), 240, 241, 251–3, 258
 - ROI (return on investment), 241, 248–51
- marketing, defined, 37–8, 239
- Marketing Leadership Council, 36
- Marketing Metrics Project, 36
- marketing performance measurement, 8, 36–7, 55–7
 - awareness of product, 42
 - balanced scorecard, 37, 48
 - branding, 46–7, 56, 240
 - cash flow, 44, 46 (*See also* discounted cash flow (DCF) metrics in marketing)
 - chain of effects in, 40–1
 - customer base, 47–50, 56
 - CRM (customer relationship management), 49–50
 - customer loyalty, 48–9
 - customer satisfaction, 47–8
 - ROC (return on customer), 240, 241, 251–3, 258
 - dashboards (*See* marketing dashboards)
 - defining marketing, 37–8
 - developing effective systems for, 51–2
 - feedback loops, 54
 - final outcomes, 43–5, 56
 - importance of, 241–2
 - intermediate outcomes, 42–3, 56
 - long-term vs. short-term measurements, 55
 - market capitalization of firm, 45
 - market share, 44
 - marketing activities (marketing mix), 41–2, 55
 - marketing assets, 45–7, 56
 - multidimensional schemes, 37
 - marketing dashboards, 240, 242–4
 - reconciling, 54–5
 - understanding, 52–3
 - practical and theoretical issues, 50–5, 57
 - productivity
 - chain of effects vs., 40–1
 - current importance of topic, 239
 - historical approaches to, 38–9
 - ROI, 39–40
 - profitability, 44, 56
 - ROC (return on customer), 240, 241, 251–3, 258
 - ROI (return on investment), 39–40, 240, 241, 248–51
 - sales, 43–5, 56
 - short-term vs. long-term measurements, 55
 - stock market performance
 - branding, 47
 - market capitalization of firm, 45
 - marketing dashboards and, 239, 247
 - reactions to marketing events, 45
 - viewed as ultimate purpose of marketing, 239
 - theoretical and practical issues, 50–5, 57
 - what to measure, 41
- Marketing Science Institute, 36, 241
- Markides, C., 272
- Marley, A., 226
- Mason, R., 224
- mathematical validity of measurement
 - frameworks. *See* measurement theory
- Mazvancheryl, S.K., 40
- McGregor, D.M., 348, 351, 456, 457
- measurability
 - in observation of knowledge work, 287, 288, 289
 - of primary school teacher performance, 352–3
- measure fixation, 422, 424
- measurement
 - informational, 297–8
 - motivational, 297–8
 - of performance generally (*See* business performance measurement)
- measurement theory, 142, 218–22
 - aspects of object to be measured, selecting, 223
 - boundaries of object to be measured, delineating, 223
 - common mistakes in business measurement systems, 220–1
 - data used to operate system, 230
 - defining measurement for purposes of, 224–32
 - fundamental concept of, 219
 - historical development of, 224–5
 - indicators vs. measurement systems, 219
 - psychology and, 225, 226
 - reasons for employing BPM, 221
 - representational or axiomatic, 225–6
 - scales, 230, 231
 - sources of data to be measured, 230
 - stakeholders, 223–4
 - steps in constructing practical measurement systems according to, 227–30
 - use of results, 232–3
 - value measurement, 218, 219, 226–7, 232–3
 - what to measure, 222–4
- measures vs. indicators, 128, 219, 220, 222
- Medicare and Medicaid, US, 394–7, 402–3
- Menon, A., 53
- methodologies. *See* frameworks and methodologies
- Meyer, Marshall, xv, 8, 113, 454
- Michelin, 266
- Milgrom, P., 275, 285, 290, 291, 293

- Miller, D., 411
- Miller, J.G., 71
- Mintzberg, Henry, 165, 168, 427
- misinterpretation, 423, 424
- misrepresentation, 422, 424
- Mitchell, Falconer, xv, 142, 179
- Mitchell, R., 223
- Mitroff, I., 224
- Mittal, V., 39–52
- Monden, Y., 67
- Moos, L., 339
- Morgan, N.A., 39–52
- motivation, 290–3
 - appropriate motivation for task, providing, 442–4
 - challenging objectives, problems associated with setting, 344–5
 - crowding out theory of, 436–7, 457
 - empirical evidence of, 438–40
 - theoretical bases for, 437–8
 - extrinsic, 434–6, 442
 - fairness, perception of, 438, 444
 - financial BPM and (*See under* financial performance measures)
 - ICs (interactive control systems), 292–3
 - intrinsic (*See under* pay for performance and motivation)
 - knowledge work measurement, 280, 290–4
 - pay for performance and (*See* pay for performance and motivation)
 - risk and individual reward incentives, 275–7
 - “theory X” assumption of, 351
 - “theory Y” assumption of, 348
- motivational measurement, 297–8
- M’Pherson, P., 222
- multi-attribute value theory (MAVT), 227
- multidimensionality
 - context dimensions of performance measurement systems and benefits of loose coupling, 478–80, 488, 489
 - of health care systems, 383
 - as key characteristic of frameworks and methodologies, 150
 - marketing dashboards and, 240, 242–4
 - in marketing performance measurement (*See under* marketing performance measurement)
 - in operations performance objectives, 68
 - of performance prism, 156
 - in primary school teacher performance measurement, 353–4
 - “silver metric”, impracticality of, 240, 243, 251, 257
 - multiple task problem, 440
- Murphy, K.J., 451, 452, 462
- myopia, 422, 424
- Narayanan, V.G., 296, 454, 465
- National Health Service (NHS).
 - See under* United Kingdom
- National Policing Plan, police force performance measurement, 365, 375
- Neely, Andy, xv, 1, 8, 64, 141, 143
- neighbourhood policing, 369, 376
- net present value (NPV), 240, 245–6, 251
- new public management (NPM), 339, 365, 413, 417, 433. *See also* public sector performance measurement
- Newtonian view of systems, 164
- NHS (National Health Service). *See under* United Kingdom
- Nokia, 177
- non-financial performance measures
 - management requirements for optimal performance measures, 115
 - problems of, 116
 - rise of, 11, 32, 114
- Noordegraaf, M., 415–16, 428
- Nørreklit, Hanne, xv, 142, 179
- Nørreklit, Lennart, xv, 142, 179
- Norton, D.P., 350. *See also* balanced scorecard approach
- NPM (new public management), 339, 365, 413, 417, 433. *See also* public sector performance measurement
- NPV (net present value), 240, 245–6, 251
- objectives, management by, 342
- observability and BPM, 280, 281–90
 - attributability, 287, 288, 289
 - context insensitivity, 289
 - evaluability, 288, 289
 - ex post* causal ambiguity, 290
 - information asymmetries, 285–90
 - inseparability, 289
 - knowledge work, 285–90
 - measurability, 287, 288, 289
 - primary school teachers, 352–3
 - problems generally arising with, 281–5
 - R–H (Ross–Holmström) model, 281–2, 288
 - reliability, 289
- Olsen, J., 452
- Olsen, S., 242
- operational responses and performance metrics, link between, 478–9
- operations management, 8, 64–5
 - balanced sets of measures for, 73, 74

- benchmarking, 74
- cost, 71–2
 - of quality, 67
 - trade-offs of, 69
- cost centre, operations viewed as, 65–6
- dependability and, 68
- five operations performance objectives, 68–9
- flexibility, 68, 72
- frameworks and methodologies, need for, 72–4
- future research agendas, 76–7
- Japan's emergence as economic power, effects of, 67
- JIT (just-in-time), 65, 71
- kaizen, 67, 69
- multidimensionality of, 68
- organizational boundaries, moving outside, 74–5
- performance measurement vs. performance management, 75–6
- productivity, 65–6
- psychology of, 76
- quality in (*See* quality)
- quantification and its unanticipated consequences, 73, 74
- speed
 - internal and external values associated with, 69
 - multidimensionality of, 68
 - trade-offs of, 69
- strategic changes in, 1960s through 1980s, 66–7
- supply chains, 75, 106–7
- trade-offs between performance objectives, 69
- opportunism, principles of BPM encouraging, 471
- Orbell, J.M., 468
- order fulfilment process, 88, 107
- organizations
 - company coherence of BPM systems, 204–5
 - coordination of actions across, 173–4
 - firm-centric performance measurement, problem of, 116–21
 - measurement across (*See* cross-organizational measurement)
 - performance defined from inside or outside, 132–3
 - radicality of innovation in terms of, 304
 - risk in assessing overall performance of, 265–9
 - subunits of
 - coordination of actions across, 173–4
 - performance responsibility, 133
- ossification, 423
- Osterloh, Margit, xv, 36, 431
- Otley, David, xv, 7, 11
- Ouchi, W.G., 291, 455
- outcome vs. process measures in health care system composite indicators, 387
- P&L (profit and loss) statements. *See under* supply chain management
- parental involvement in primary school teacher performance measurement, 354
- patent citation analysis. *See under* innovation performance measurement
- Paton, W.A., 210
- pay for performance and motivation, 431, 433, 444
 - appropriate motivation for task, providing, 442–4
 - clear link between performance and reward as basic principle, 455–7
 - complex tasks and, 434
 - crowding out theory, 436–7, 457
 - empirical evidence of, 438–40
 - theoretical bases for, 437–8
 - extrinsic motivation, 434–6, 442
 - fairness, perception of, 438, 444
 - incomplete contracts, situations involving, 433, 441, 442
 - intrinsic motivation, 434–6, 456–7
 - crowding out of (*See* subhead “crowding out theory”, *this entry*)
 - disadvantages of, 441–2
 - factors favouring or undermining, 443–4
 - situations requiring, 440–1, 442
 - in knowledge work, 434
 - perversity and dysfunction related to, 433–4, 456
 - for primary school teachers, 351–2
 - conflicts between objectives, 356–8
 - subjectivity, reliability and measurability problems, 352–3
- PCA (principal components analysis), 388
- peer-based performance reviews, 176–7
- Peppers, D., 240, 251, 252, 258
- performance-based pay. *See* pay for performance and motivation
- Performance Management in Schools initiative, 340. *See also* primary school teachers, measuring performance of
- performance measurement, generally. *See* business performance measurement
- performance measurement matrix (PMM), 145, 149, 150
- performance prism, 141, 143–4, 151–6
 - capabilities, 155
 - case studies, 156–60
 - common findings of, 160

- performance prism, (cont.)
- DHL UK, 156–8
 - private sector Canadian firm, 159
 - private sector Middle Eastern firm, 159–60
 - public sector UK health care organization, 158
 - public sector US employment services, 159
 - small charitable association, UK, 158
 - competition in, 151, 155
 - key characteristics of existing frameworks, incorporating, 149–51, 156
 - managerial buy-in, 160
 - primary school teacher performance measurement using, 353
 - processes, 154
 - reasons for adopting, 160
 - stakeholder approach, 151–3
 - strategies, 153–4
 - timescale of projects, 160
- perversity and dysfunction associated with basic BPM principles
- accurate measurement of performance, 463, 464, 465–6
 - clear specification of performance criteria, 462
 - opportunism, principles encouraging, 471
 - reward and performance, clear link between, 456
- perversity and dysfunction in airline delay tracking and assignment systems, 459–61
- perversity and dysfunction in pay for performance, 433–4, 456
- perversity and dysfunction in public sector
- performance measurement, 238, 417–18
 - anomalous results, 427–8
 - cybernetics, 426
 - different types of, situations leading to, 422–6
 - gaming, 423, 424
 - measure fixation, 422, 424
 - misinterpretation, 423, 424
 - misrepresentation, 422, 424
 - myopia, 422, 424
 - ossification, 423
 - police force, 366
 - pupils' academic progress as teacher performance measure, 346
 - reasons for employing measurement associated with particular perversities, 423–6
 - suboptimization, 422
 - tunnel vision, 422
- Pfanzagel, J., 226
- phenomenology, 189
- facts, truth's phenomenological grounding in, 187–8
 - performance creating firm as phenomenon, 198–9
- philosophical issues in BPM. *See* theoretical foundations of BPM
- Pidd, Mike, xv, 238, 408, 418, 419
- Piercy, N., 45
- Pike, Stephen, xv, 142, 218, 222, 232
- Pittman, T.S., 436
- plan–do–check–act (PDCA) cycle, 168, 326, 409
- planning
- in adaptive organizations (*See* adaptive planning model)
 - marketing dashboard metrics for, 253–6, 257
 - SCM and, 90
- Plato's cave analogy, 129
- Plunkett, J.J., 70
- PMM (performance measurement matrix), 145, 149, 150
- Poisson regression analysis, 312, 313
- police force performance measurement, 237, 363
- as accounting system, 372–3
 - antisocial behaviour, 369
 - BCU (basic command unit), role of, 367, 368, 373
 - bio power, 371, 375
 - case studies of four police forces, 366, 369
 - citizen focus, 369, 376
 - Compstat model, 367
 - development of, 365–6
 - discipline, 370–1, 372, 374–5, 376–9
 - fear of crime vs. level of crime, 369
 - funding and performance, links between, 375
 - governmentality, 371–2, 375–6, 376–9
 - history, structure and governance of police force, 363–5
 - iQuanta system, 367, 374, 380
 - as knowledge and power, 363, 369–72, 379–80
 - measurement to management, shift between, 376
 - National Policing Plan, 365, 375
 - neighbourhood policing, 369, 376
 - perverse and dysfunctional consequences of, 366
 - Police Standards Unit, 367, 368, 373, 375
 - political dimensions of, 363, 364, 366, 373, 375, 380
 - PPAF (Policing Performance Assessment Framework), 365, 374, 375, 380
 - PSAs (public service agreements), 365, 367, 373, 375
 - self-assessment, 371, 375
 - sovereignty, 370–1, 372, 373–4, 376–9
 - SP's (statutory performance indicators), 365, 375

- transitions within sovereignty–discipline–governmentality triangle, 376–9
- political dimensions
 - of police force performance measurement, 363, 364, 366, 373, 375, 380
 - of public sector performance measurement generally, 416–17, 421, 426
- Pollitt, C., 408
- Popper, Karl, 214
- Porras, J.I., 243, 272
- Porter, Michael, 167
- positivism, 189
- Powell, Tony, 238, 318
- power
 - police force performance measurement as knowledge and, 363, 369–72, 379–80
 - SCM (supply chain management), power and leadership structure, 91
- Power, M., 73
- PPAF (Policing Performance Assessment Framework), 365, 374, 375, 380
- practical applications and challenges, 4, 237–8
- pragmatic constructivism, 142. *See also* theoretical foundations of BPM
- coherence
 - communication and social coherence, 196
 - pragmatic coherence to purpose and objectives of BPM systems, 198
- facts, 181
- integrated learning theory (*See* integrated learning theory of truth)
- mechanical realism vs., 185–6
- reality, concept of, 181
- truth, 186, 196–7, 212–14
- Prendergast, C., 435, 456
- primary school teachers, measuring performance
 - of, 237, 339–40, 359
 - agency theory, 350
 - conflicts between objectives, 356–8
 - “customer” (parent/pupil) involvement in, 354
 - methodology of studying, 340
 - multiple appraisers, 353–4
 - objectives of performance measurement
 - conflicts between, 356–8
 - defined and described, 341–2
 - objectives of teachers and goal-setting theory, 341–2
 - challenging objectives, problems associated with setting, 344–5
 - conflicts between performance measurement objectives and, 357
 - defining “performance” and “objectives”, 342–4
 - environmental uncertainty and control over performance, 345–6
 - focusing attention on, measurement to achieve, 342
 - strategy implementation through goal alignment, 349–51
- overall performance measurement, 354–6
- pay rates and, 351–2
 - conflicts between objectives, 356–8
 - subjectivity, reliability and measurability problems, 352–3
- Performance Management in Schools initiative, 340
- Performance Prism framework used for, 353
- policy implications, 359
- for professional development, 348
 - availability issues, 349
 - conflicts with other objectives, 356–7, 358
 - financing issues, 349
 - identification and communication of needs regarding, 348
- pupils’ academic progress as teacher performance measure, 346
 - availability of information, 347–8
 - conflicts between objectives, 357
 - unintended/dysfunctional consequences of, 346
 - validity of, 347
- self-appraisal, 353
- stakeholders, 350, 353–4
- standards for teachers, 354–6
- for strategy implementation through goal alignment, 349–51, 358
- subjectivity, reliability and measurability problems, 352–3
- subordinate involvement in, 354
- 360-degree feedback, 353
- principal components analysis (PCA), 388
- principle agent theory, 435
- processes
 - business process redesign/re-engineering, 151, 154
 - health care system composite indicators, using process vs. outcome measures in, 387
 - Macro Process Model of the organization, 146
 - performance prism and, 154
 - SCM (supply chain management)
 - business or management processes, 86–90, 105–6
 - process-based view of, 83
 - statistical process control, 70
 - product and service agreements (PSAs)
 - CRM (customer relationship management) and, 87
 - customer service management and, 87
 - supplier relationship management, 89

- product development and commercialization, 89
- product flow facility structure, 91
- production, generally. *See* operations management
- productivity
- in marketing (*See under* marketing performance measurement)
 - in operations management, 65–6
 - in public sector performance measurement, 408
- professional development of primary school teachers. *See under* primary school teachers, measuring performance of
- professionals
- creative work, intrinsic motivation required for, 440
 - knowledge workers (*See* knowledge work measurement)
 - in public sector performance measurement, 415–16
 - teachers (*See* primary school teachers, measuring performance of)
- profit and loss (P&L) statements. *See under* supply chain management
- profitability
- as financial performance measurement, 13, 15, 17, 18, 19
 - as marketing performance measurement, 44, 56
- project management, CbM in. *See* context-based measurement
- promotions, monitoring, 254
- PSAs (product and service agreements). *See* product and service agreements
- PSAs (public service agreements) and police force performance measurement, 365, 367, 373, 375
- psychological contract theory, 437–8
- psychology
- intrinsic motivation in, 435
 - measurement theory and, 225, 226
 - observability problem, 283
 - of operations management, 76
 - performance prism and, 152
 - TASK (talent, skill and knowledge) differentials, 295
- psychometrics, 226
- public sector performance measurement, 4, 237–8, 383–4
- ambiguity and uncertainty, 415–16, 426
 - Atkinson Review, 408, 409
 - cybernetics, 413–15, 426
 - education (*See* primary school teachers, measuring performance of)
 - health care (*See* health care systems, use of composite indicators in)
 - NPM (new public management), 339, 365, 413, 417, 433
 - pay for performance in, 433 (*See also under* primary school teachers, measuring performance of)
 - perversity and dysfunction (*See* perversity and dysfunction in public sector performance measurement)
 - police force (*See* police force performance measurement)
 - productivity, 408
 - professionally skilled workers, 415–16
 - professionals in, 415–16
 - reasons for employing, 409–17
 - central control, 413–16, 419, 421, 426
 - competences, identifying, 410–11, 419, 424
 - determining what works, 409–10, 419, 421, 424
 - perversity and dysfunction associated with, 423–6
 - public accountability, 411–13, 419, 424
 - SSM (soft systems methodology) used to examine, 418–21
 - symbolic/political reasons, 416–17, 421, 426
 - RSS (Royal Statistical Society) review, 409, 423
- public service agreements (PSAs) and police force performance measurement, 365, 367, 373, 375
- pupil involvement in primary school teacher performance measurement, 354
- pupil progress as teacher performance measure. *See under* primary school teachers, measuring performance of
- pyramid of ratios, 14–17, 144
- quality, 70–1
 - cost of, 67
 - EFQM (*See* European Foundation for Quality Management)
 - EOQ (economic order quality) model, 67, 70
 - internal and external values associated with, 69
 - Malcolm Baldrige National Quality Award, 1, 70, 149
 - multidimensionality of, 68
 - rise in popularity of “quality gurus”, 67
 - statistical process control, 67
 - TQM (total quality management), 70, 151, 409, 418
 - trade-offs of, 69
- Quinn, J.B., 256
- radical innovation. *See under* innovation performance measurement
- R&D (research and development) collaboration and innovation performance. *See* innovation performance measurement

- Rastas, T., 222
- rationalism, 192
- ratios, pyramid of, 14–17, 144
- reality/realism
- communication and, 183–4
 - concept of, 181
 - integration and, 184
 - mechanical realism vs. pragmatic constructivism, 185–6
 - misleading use of, 214–15
- reasons for employing BPM, 160, 221. *See also* *under* public sector performance measurement
- Rego, L.L., 40
- regulations and laws
- risk management, 262
 - stakeholders, legal and regulatory communities as, 152
- Reichheld, F.E., 49
- Reinartz, W., 39
- relative performance contracts, 174–8
- relativity of concept of performance, 135–6
- reliability
- in observation of knowledge work, 289
 - performance-related pay for primary school teachers and, 352–3
- remuneration committees (RemCos), 276
- reporting on KPIs (key performance indicators), 170–1
- representational or axiomatic measurement theory, 225–6
- research and development (R&D) collaboration and innovation performance. *See under* innovation performance measurement
- residual income, 25
- resource management, 171–3, 393
- responsibility and accountability
- airline delay tracking and assignment, accepting responsibility in, 465–6
 - financial performance measures, responsibility centres for, 23
 - marketing accountability, 242
 - public accountability, 411–13, 419, 424
 - subunit performance, 133
 - teams, focus of accountability on, 174–8
- restaurant chain case study. *See under* loose coupling in performance measurement systems
- results–determinants framework, 146, 147, 148
- return on customer (ROC), 240, 241, 251–3, 258
- return on investment (ROI)
- financial performance measurement and, 14, 18–22
 - in marketing, 39–40, 240, 241, 248–51
- returns management, 89
- returns, shareholder rate of, 18–22
- revenue centres, 23
- rewards
- clear link between performance and, 455–7, 466–8
 - pay for performance (*See* pay for performance and motivation)
 - risk and, 91, 275–7
- Ridgway, V.F., 73, 77, 284, 417
- Riley, F.D.O., 47
- risk and BPM, 237, 261–4, 277
- definition of risk, 264
 - economic profit models, 266, 273, 274
 - “face” and risk appetite, 264
 - in health care systems, 392
 - heat charts, 270
 - internal risk management, 269–75
 - legal and regulatory requirements, 262, 263
 - lender ratings, 268–9
 - overall company performance, assessment of, 265–9
 - perspectives on links between, 264, 265, 277
 - for quoted companies, 266–8
 - reward and risk, 91, 275–7
 - SCM (supply chain management), risk and reward structure in, 91
 - for unquoted companies, 268
 - VaR (value at risk), 271, 274
- Roberts, John, xvi, 237, 239, 275, 285
- ROC (return on customer), 240, 241, 251–3, 258
- Rogers, M., 240, 251, 252, 258
- ROI. *See* return on investment
- rolling forecast, 169–70, 172
- ROMI (return on marketing investment), 39–40, 240
- Rook, P., 320
- Roos, Goran, xvi, 142, 218, 222
- root definitions, 418–19, 421
- Rosenbloom, R.S., 307
- Ross, S.A., 281
- Royal Society of Arts, Manufacturers and Commerce, 2
- Royal Statistical Society (RSS) review of public sector performance measurement systems, 409, 423
- Ruch, W.A., 66
- Rudduck, Jean, 354
- Rust, R.T., 251
- Ryan, R.M., 438
- sales as marketing performance measurement, 43–5, 56
- Sarbanes–Oxley Act, 17

- SAST (strategic assumption surfacing and testing), 224
- satisfaction of customers, 47–8
- scales
- composite indicators, developing, 388–9
 - in measurement theory, 230, 231
- Scholes, J., 418
- Schonberger, R.J., 67
- SCM. *See* supply chain management
- Searle, J., 206, 209
- SEC (Securities and Exchange Commission), 262, 263
- Securities and Exchange Commission (SEC), 262, 263
- self-appraisal
- by police force, 371, 375
 - by primary school teachers, 353
- Selznick, P., 452, 462
- “semi-confusing” information systems, 481
- senior management support for BPM projects, 160
- “sense and respond” management, shift from “make and sell” paradigm to, 163–4
- sensitivity analysis for composite indicators in health care systems, 393
- Sevin, Charles, 39
- shareholders
- financial reporting to, 18–22
 - individual reward incentives and risk, 275–7
 - marketing dashboards and shareholder value, 239, 247
 - risk in BPM calculated using TSR (total shareholder return), 266–8
- Sharpe ratio, 267
- Shell, 2
- Shervani, T.A., 44, 45
- Sheth, J.N., 39–52
- “silver metric”, impracticality of, 240, 243, 251, 257
- Simon, C.J., 46
- Simon, H.A., 456, 462
- Simons, R., 292
- Sisodia, R.S., 39–52
- Skandia, 2
- skilled workers
- creative work, intrinsic motivation required for, 440
 - knowledge workers (*See* knowledge work measurement)
 - in public sector performance measurement, 415–16
 - teachers (*See* primary school teachers, measuring performance of)
- Skinner, Wickham, 64, 67, 72
- Slack, N., 72
- SMART (strategic measurement and reporting technique) pyramid, 145, 146, 150
- Smith, Peter, xvi, 237, 383, 417, 422, 423, 426
- social constructivism, 180, 196
- soft systems methodology (SSM), 418–21
- Sonnenfeld, J., 223
- sources of data to be measured, 230
- Southwest Airlines, 171, 173, 177, 178
- sovereignty and police force performance
- measurement, 370–1, 372, 373–4, 376–9
- speed in operations management
- internal and external values associated with, 69
 - multidimensionality of, 68
 - trade-offs of, 69
- SPIs (statutory performance indicators), police force performance measurement, 365, 375
- Srivastava, R.K., 44, 45
- SSM (soft systems methodology), 418–21
- stakeholders
- measurement theory and, 223–4
 - in performance prism framework, 151–3
 - in primary school teacher performance measurement, 350, 353–4
- star ratings, 401–2, 424
- statistical process control, 70
- statutory performance indicators (SPIs), police force performance measurement, 365, 375
- Sterling, R.R., 210
- Stern Stewart, 11, 25, 30, 33
- Stevens, S.S., 225
- stock market performance
- as marketing performance measurement (*See under* marketing performance measurement)
 - risk in BPM and, 266–8
- stockholders. *See* shareholders
- strategic assumption surfacing and testing (SAST), 224
- strategic measurement and reporting technique (SMART) pyramid, 145, 146, 150
- strategy
- in adaptive organizations, 165–7
 - mapping, 34
 - Mintzberg’s description of, 165, 166
 - in performance prism framework, 153–4
 - primary school teacher performance measurement for implementation of, 349–51, 358
- student involvement in primary school teacher performance measurement, 354
- student progress as teacher performance measure. *See under* primary school teachers, measuring performance of
- suboptimization, 422

- subunits
 - coordination of actions across, 173–4
 - performance responsibility, 133
- subjectivity problems of performance-related pay
 - for primary school teachers, 352–3
- subordinate involvement in performance
 - measurement, 354
- Sullivan, M.W., 46
- Suppes, P., 225
- supply chain management (SCM), 8, 82–4
 - alignment of performance measures across supply chain, 97–9, 99–100
 - analysing links in supply chain, 102–3
 - business or management processes, 86–90, 105–6
 - conceptual framework for, 84–5
 - corporate and supply chain performance, linking, 97
 - corporate culture, 91
 - cross-organizational comparison and replication, 107–9
 - customer base
 - CRM (customer relationship management), 87, 102–3
 - customer service management, 87
 - demand management, 88
 - order fulfilment, 88, 107
 - performance of customers, 100
 - P&L statements (*See* subhead “P&L (profit and loss) statements”, *this entry*)
 - product development and commercialization, 89
 - returns management, 89
 - defining, 95
 - demand management, 88
 - EVA[®] (economic value added) and, 102, 103, 104, 106–7
 - frameworks
 - GSCF conceptual framework, 84–5
 - metrics, framework for developing, 100–9
 - future research directions, 109–10
 - information flow structure, 91
 - inventory turns, 92–5
 - management components, 90–1
 - management methods and philosophies, 91
 - manufacturing flow management, 88
 - mapping supply chain, 101, 102
 - metrics of, 83–4
 - framework for developing, 100–9
 - need for new types of, 95–7
 - problems with existing, 92–5
 - strategy and metrics, linking, 97–9, 99–100
 - network structure, 85–6
 - operations management and, 75, 106–7
 - order fulfilment, 88, 107
 - P&L (profit and loss) statements, 105
 - business or management processes, realignment of, 105–6
 - developing, 103–5
 - operational measures aligned with, 106–7
 - value of, 110
 - planning and control of operations, 90
 - power and leadership structure, 91
 - process-based view of, 83
 - product development and commercialization, 89
 - product flow facility structure, 91
 - PSAs (product and service agreements)
 - CRM (customer relationship management) and, 87
 - customer service management and, 87
 - supplier relationship management, 89
 - replication across organizations, 107–9
 - returns management, 89
 - risk and reward structure, 91
 - strategy and metrics, linking, 97–9, 99–100
 - suppliers
 - performance of, 100
 - P&L statements (*See* subhead “P&L (profit and loss) statements”, *this entry*)
 - SMI (supplier managed inventory), 102, 105
 - SRM (supplier relationship management), 89, 102–3
 - time compression, 96
 - work structure, 91
- symbolic reasons for public sector performance
 - measurement, 416–17, 421, 426
- Szymanski, D.M., 48
- Tableau de Bord, 148, 150
- tacit knowledge, intrinsic motivation required for transfers of, 441
- Tang, S.H., 438
- TASK (talent, skill and knowledge) differentials, 238, 280–1, 295–8, 299
- Taylor, F.W., 456
- teaching, BPM in. *See* primary school teachers, measuring performance of
- teams, focus of accountability on, 174–8
- technologically radical innovation, 305
- theoretical foundations of BPM, 4, 7–9, 142, 179–80. *See also* coherence theory; integrated learning theory of truth; measurement theory; pragmatic constructivism
- communication, 172, 183–4
- correspondence theory, 188–9
- dimensionalization of measurement, 191, 199–200, 212

- theoretical foundations of BPM, (cont.)
- epistemic platform, 189–91
 - data, truth of vs. belief in, 208–12
 - levels of, 201
 - modelling and dimensionalizing, 199–200, 212
 - facts, 181
 - correspondence and, 188–9
 - in mechanical realism, 185–6
 - truth's phenomenological grounding in, 187–8
 - grounded theory, 189, 190, 191, 192, 199–200, 212
 - integration, 184, 196–7
 - learning approach to validity and truth, *See also* integrated learning theory of truth)
 - logic, 182, 192–3
 - marketing performance measurement, 50–5, 57
 - phenomenology, 189
 - facts, truth's phenomenological grounding in, 187–8
 - performance creating firm as phenomenon, 198–9
 - positivism, 189
 - rationalism, 192
 - reality, concept of, 192 (*See also* reality/realism)
 - social constructivism, 180, 196
 - truth, 142, 179–80, 214–15
 - belief vs., 208–12
 - coherence theory of, 192–3
 - correspondence theory of, 188–9
 - integrated learning approach to, 196–7 (*See also* integrated learning theory of truth)
 - phenomenological grounding of, 187–8
 - pragmatic perspective on, 186, 196–7, 212–14
 - proactive vs. realized or pragmatic, 196–7, 212–14
 - theoretical conditions for valid truth in BPM, 186
 - values, 182–3
 - belief vs. truth, 208–12
 - coherence theory and, 193–4
 - “theory X” assumption of motivation in organizations, 351
 - “theory Y” assumption of motivation in organizations, 348
 - Thomas, A.L., 210
 - Thompson, J.D., 291, 455
 - 360-degree feedback, 353
 - time compression, 96
 - timescale of BPM projects, 160
 - Titmuss, R.M., 439
 - Tobin's q, 47
 - Tomkins, 178
 - Torrington, D., 342, 356
 - total quality management (TQM), 70, 151, 409, 418
 - total shareholder return (TSR) and risk in BPM, 266–8
 - Toyota, 67, 163, 165, 172–3
 - TQM (total quality management), 70, 151, 409, 418
 - Trajtenberg, M., 308
 - transaction cost theory, 435
 - transfer price, 24
 - Triands, H.C., 264
 - truth. *See under* theoretical foundations of BPM
 - TSR (total shareholder return) and risk in BPM, 266–8
 - tunnel vision, 422
 - Turnbull requirements, 262
 - Turney, P.B.B., 65
 - Tushman, M., 307
 - Tversky, A., 227
 - uncertainty. *See* ambiguity and uncertainty; control issues; environmental factors
 - United Kingdom
 - NHS (National Health Service)
 - composite indicators used to evaluate, 398–9, 401–2
 - performance prism case study, 158
 - perversity and dysfunction in, 424
 - public accountability of, 411, 424
 - performance prism case studies (*See under* performance prism)
 - police force (*See* police force performance measurement)
 - primary school teachers (*See* primary school teachers, measuring performance of)
 - United States
 - Medicare and Medicaid, 394–7, 402–3
 - performance prism framework used in public sector employment services, 159
 - SEC (Securities and Exchange Commission), 262, 263
 - user-radical innovation, 304
 - valid measurement frameworks
 - applied mathematics and (*See* measurement theory)
 - in integrated learning theory (*See* integrated learning theory of truth)
 - value at risk (VaR), 271, 274
 - value measurement
 - in combining variables for composite indicators, 389–90
 - EVM (earned value measurement), 320
 - in measurement theory, 218, 219, 226–7, 232–3

- values in BPM, 182–3
 - belief vs. truth, 208–12
 - coherence theory, 193–4
- van de Kragt, A.J.C., 468
- Varadarajan, P.R., 53
- variable pay for performance. *See* pay for performance and motivation
- Vickers, J., 454
- Vollmann, T.E., 71

- Wallerand, Jan, 178
- Walton, R., 473
- Wang Laboratories, 145
- WBS (work breakdown structure), 319–21

- Weiss, A., 309, 313
- Weltanschauung, 419
- WHO (World Health Organization),
 - use of composite indicator by, 399, 401
- Wiersma, U.J., 438
- Williams, R.S., 342, 350, 353, 356
- Williamson, O., 451
- Wivel, Teddy, 151
- Womack, J.P., 75
- work breakdown structure (WBS), 319–21
- World Health Organization (WHO), use of
 - composite indicator by, 399, 401
- Worldcom, 179
- Wragg, E.C., 352