Psychometrics

An Introduction

Second Edition

R. Michael Furr
Wake Forest University

Verne R. Bacharach
Appalachian State University

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

CHAPTER 4

Test Dimensionality and Factor Analysis

Imagine that a colleague wishes to use a personality inventory that includes the following six adjectives: *talkative*, *assertive*, *imaginative*, *creative*, *outgoing*, and *intellectual*. For this brief questionnaire, participants are asked to consider the degree to which each adjective describes their personality in general. Your colleague asks for your opinion of this common, adjective-based form of personality assessment. You consider the inventory for a moment, and you begin to wonder—what exactly does this inventory measure? Does it measure six separate facets of personality, with each facet being reflected by a single adjective? Or does it measure a single construct? If so, then what is that construct—what do these six adjectives have in common as a psychological characteristic or dimension? Or are there two or three separate dimensions reflected within these six adjectives? How will this questionnaire be scored?

Take a moment to think about the six adjectives on the short inventory, and group them into clusters that seem to share some common meaning. That is, group them in terms of their similarity to each other. Some people might suggest that the questionnaire includes only two sets of items. For example, some might argue that talkative, assertive, and outgoing are three variations on one attribute (let us call it "extraversion") and that imaginative, creative, and intellectual are three variations on another attribute (let us call it "openness to experience"). From this perspective, responses to these six personality adjectives reflect two basic dimensions: one set of responses that are a function of an extraversion dimension and one set of responses that are the result of an openness-to-experience dimension.

In contrast, some people might suggest that the six adjectives reflect three dimensions, not two. Specifically, "talkative," "assertive," and "outgoing" might go together, and "imaginative" and "creative" might go together, but "intellectual" is importantly different from the other five items. From this perspective, responses to the six items reflect three basic dimensions. Put another way, these six test items essentially reflect three ways in which people differ from each other psychologically.

This example illustrates the issue of test dimensionality, which is a fundamental consideration in test development, evaluation, and use. There are at least three fundamental psychometric questions regarding the dimensionality of a test, and the answers to these questions have important implications for evaluating the psychometric properties of any behavioral test, for appropriately scoring on a test, and for the proper interpretation of test scores.

In this chapter, we discuss the concept of dimensionality, the key questions related to dimensionality, and the implications that dimensionality has for test construction, evaluation, use, and interpretation. Indeed, as shown in Figure 4.1, the answers to the three key questions lead to three main types of tests: (1) unidimensional tests, (2) multidimensional tests with correlated dimensions, and (3) multidimensional tests with uncorrelated dimensions. Test developers and test users must understand which type of test is being developed or used, because these tests have important psychometric differences from each other.

Given the importance of understanding a test's dimensionality, we also describe one way the dimensionality questions can be answered quantitatively. We describe the way in which test developers, test evaluators, and test users identify the number of dimensions reflected by a test, the meaning of those dimensions, and the degree

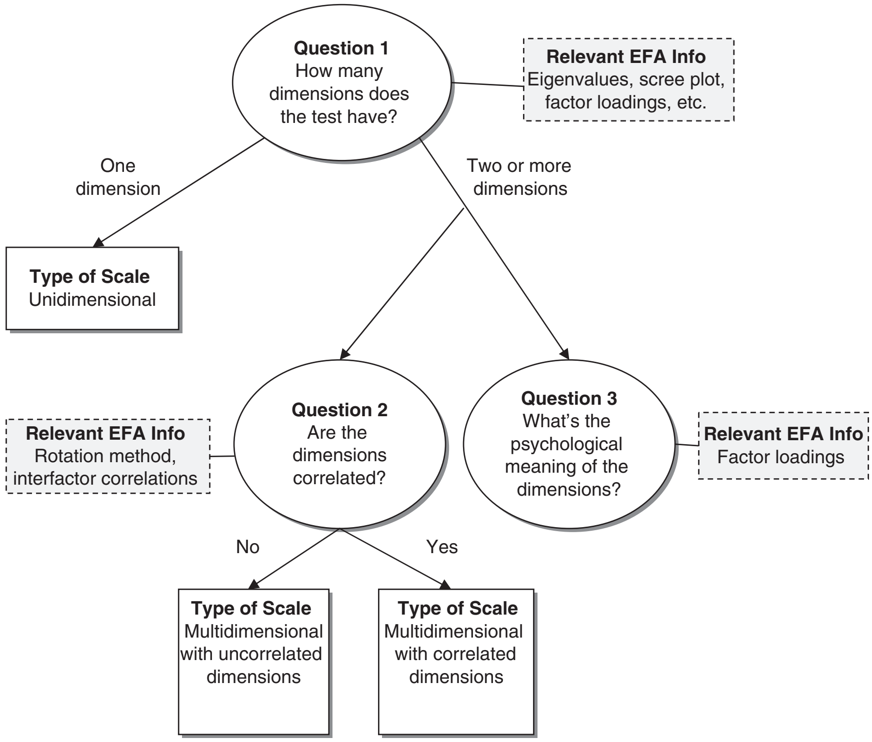


Figure 4.1 Three Core Questions of Dimensionality, Three Types of Tests, and the Relevant Information From Exploratory Factor Analysis (EFA)

to which the dimensions are associated with each other. A statistical procedure called *factor analysis* is an extremely useful tool in the psychometrician's statistical toolbox. Although factor analysis can be a highly technical procedure, we will present its general logic and use in a way that is accessible to those who do not have a great interest or background in advanced statistics. A basic understanding of factor analysis can provide a solid foundation for several important psychometric issues.

Test Dimensionality

If you step on a bathroom scale, the resulting number on the scale is a value that tells you something about one of your physical attributes or features—your weight. As a human being, you have many other physical attributes, including height, skin color, length of hair, and so on. When you weigh yourself, the number that represents your weight should not be influenced by attributes such as your hair color, your height, or your age. The "score" on the bathroom scale should (and does) reflect one and only one physical dimension.

Similarly, if we have a psychological test that yields some kind of number, then we would like to think of the number as a value representing a single psychological feature or attribute. For example, suppose you had a test of courage. If you have a test that produces scores that can be treated as if they are real numbers, then a person's score on the test might indicate the amount of courage that he or she had when taking the test. We could then think of courage as an attribute of that person and the test score as an indication of the amount of the person's courage. The score on the courage test *should* reflect one and only one psychological dimension.

As a general rule (but not always), when we measure a physical or psychological attribute of an object or a person, we intend to measure a *single* attribute of that object or person. In the case of weight, we try to measure weight so that our measurement is not affected by other attributes of the person being measured. Furthermore, it would not be reasonable to measure someone's weight, measure his or her hair length, and then add those two scores together to form a "total" score. Clearly, the total score would be a blend of two physical dimensions that are almost totally unrelated to each other, and the combination of the two scores would have no clear interpretation. That is, the total score would not have clear reference to a single physical attribute and thus would have no clear meaning. Similarly, it would not be reasonable to measure someone's courage, measure his or her verbal skill, and then add those two scores together to form a "total" score. Again, the total score would be a blend of two dimensions that are clearly unrelated to each other (i.e., one's courage is probably unrelated to one's level of verbal skill). Combining test scores from two independent psychological attributes produces a total score that has no clear meaning.

As discussed in our presentation of composite scores, the scores from a wide variety of psychological tests are based on multiple questions or test items. For example, personality tests range in length from 5 or fewer questions to several hundred questions. In scoring such tests, item responses are combined in some way, usually by computing one or more scores of some kind, and these combined scores are used to reflect the psychological attribute(s) of interest. These scores are referred to as composite scores, and ideally, a composite score reflects one and only one dimension. However, a test may include items that reflect more than a single dimension.

Three Dimensionality Questions

As mentioned earlier, there are at least three core questions regarding a test's dimensionality. First, how many dimensions are reflected in the test items? As we shall see, some tests reflect one and only one dimension, while others reflect two or more psychological dimensions. This issue is important because each dimension of a test is likely to be scored separately, with each dimension requiring its own psychometric analysis.

The second core dimensionality question is this: If a test has more than one dimension, then are those dimensions correlated with each other? As we shall see, some tests have several dimensions that are somewhat related to each other, while other tests have several dimensions that are essentially independent. This issue is important, in part, because the nature of the associations among a test's dimensions has implications for the meaningfulness of a "total score" for a test.

Third, if a test has more than one dimension, then what *are* those dimensions? That is, what psychological attributes are reflected by the test dimensions? For example, in the six-adjective personality test described previously, does the first dimension reflect the psychological attribute of extraversion or some other attribute? The importance of this issue should be fairly clear—if we score and interpret a dimension of a test, we must understand the score's psychological meaning.

Figure 4.1 summarizes these questions and illustrates their connections to three different types of tests. These types of tests have different properties, different implications for scoring and for psychometric evaluation, and ultimately different psychological implications.

Unidimensional Tests

The first question regarding test dimensionality concerns the number of dimensions reflected in a set of test items. Some tests include items that reflect a single psychological attribute, and others include items that reflect more than one attribute.

When a psychological test includes items that reflect only a single attribute of a person, this means that responses to those items are driven only by that attribute (and, to some degree, by random measurement error—see Chapters 5–7). In such cases, we say that the test is *unidimensional*, because its items reflect only one psychological dimension.

Consider a multiple-choice geometry exam given in a classroom. Typically, a student takes the exam and receives a score based on the number of questions that he or she answers correctly. The student's score is then interpreted as a measure of the amount of his or her "knowledge of geometry." This interpretation makes sense only if the answers to all the test items truly require knowledge of geometry, and *only* knowledge of geometry. For example, if we can believe that the test does not (mistakenly) include algebra items, calculus items, or vocabulary items in addition to geometry items, then we can indeed have some confidence in interpreting test scores as reflecting knowledge of geometry. That is, we could assume that the answers to each of the questions on the test are affected by that single psychological attribute. Such a test would be thought of as unidimensional. In addition, the test items or questions would have the property of *conceptual homogeneity*—responses to each item would be a function of the same psychological attribute.

The concept of a unidimensional test is illustrated in Figure 4.2. This figure uses formatting that is standard for graphically representing a test's dimensionality (or factorial structure, as we shall describe later). In such figures, a circle or oval represents a hypothetical psychological attribute or latent variable that affects participants' responses to test questions. Returning to the geometry test example, the circle would represent "knowledge of geometry" because it is the psychological property that (supposedly) determines whether a student answers the test items correctly. Correspondingly, in figures like Figure 4.2, squares or rectangles represent responses to each of the test questions. Finally, the arrows' directionality (i.e., they point from the attribute to the responses) represents the idea that the psychological attribute affects responses to test questions. For example, they represent the assumption that knowledge of geometry (as a psychological ability) is what affects students' answers to the test questions. Because it shows a single psychological attribute affecting participants' responses, this figure illustrates a unidimensional test.

As we have mentioned, a test's dimensionality has implications for its scoring, evaluation, and use. For a unidimensional test, only a single score is computed, reflecting the single psychological attribute measured by the test. That is, all the items are combined in some way (usually through averaging, summing, or counting) to form a composite or "total" score. For example, if it is indeed unidimensional, the geometry test produces a single score (e.g., the total count of the number of correctly answered questions) reflecting "knowledge of geometry." In terms of

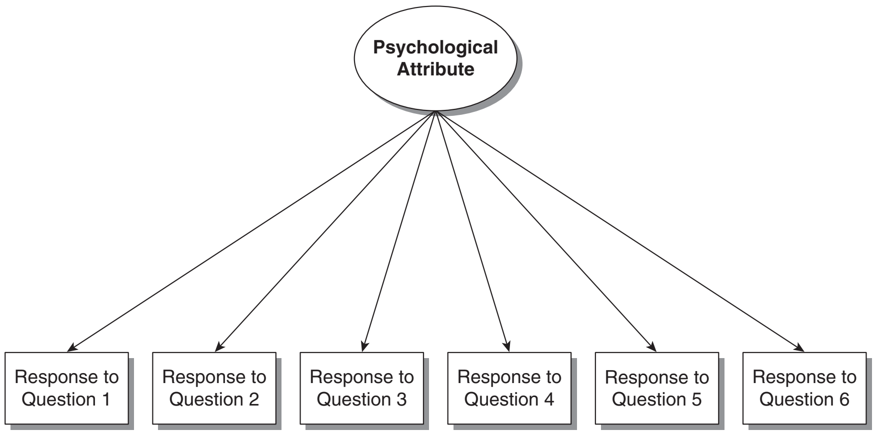


Figure 4.2 Unidimensional Test

psychometric evaluation, psychometric quality is evaluated for the single score that is obtained from a unidimensional test. In later chapters, we shall discuss reliability and validity, which reflect the psychometric quality of test scores. For unidimensional tests, reliability and validity should be estimated and evaluated for the total score produced by the test. In terms of test use, test users compute and interpret the total score produced by a unidimensional test.

Multidimensional Tests With Correlated Dimensions (Tests With Higher-Order Factors)

When a psychological test includes items reflecting more than one psychological attribute, the test is considered multidimensional. In such cases, we confront a second dimensionality question—are the test's dimensions associated with each other? As shown in Figure 4.1, the answer to this question differentiates two types of tests. When a test has multiple dimensions that are correlated with each other, the test can be considered a *multidimensional test with correlated dimensions* (this has also been called a test with higher-order factors).

Intelligence tests such as the Wechsler Intelligence Scale for Children (WISC-IV) (Wechsler, 2003a, 2003b) and the Stanford-Binet (SB5) (Roid, 2003) are examples of multidimensional tests with correlated dimensions. These tests include groups of questions that assess different psychological attributes. The groups of questions are called subtests, and they each reflect a different facet of intelligence. For example, the SB5 has five subtests: (1) one to measure fluid reasoning, (2) one to measure general knowledge, (3) one to measure quantitative processing ability, (4) one to measure visual-spatial processing ability, and (5) one that is thought to measure working memory. Research by test developers and test evaluators has shown that the subtests of the SB5 are correlated with each other. That is, a participant who scores relatively high on one subtest is likely to score relatively high on the other subtests as well.

As we have mentioned, a test's dimensionality has important implications for the scoring, evaluation, and use of the test. Multidimensional tests with correlated dimensions can produce a variety of scores. Typically, each subtest has its own subtest score. In principle, each subtest is, itself, unidimensional, and the questions in each subtest are conceptually homogeneous. For example, the quantitative processing subtest of the SB5 might require a test taker to answer 10 questions. Presumably, responses to each of those 10 questions reflect only quantitative processing and not one of the constructs represented by the other subtests. That is, a person's responses to the 10 questions are affected only by the person's quantitative processing skills and not some other psychological attribute. If a subtest is unidimensional, then the subtest's score is interpretable with regard to a single psychological attribute.

In addition to scores for each subtest, multidimensional tests with correlated dimensions are often scored in a way that produces a total score, combined across several subtests. That is, subtest scores are often combined with each other (again, either through summing or by averaging the scores) to produce a *total test score.* For example, the five subtest scores from the SB5 are combined to form an overall

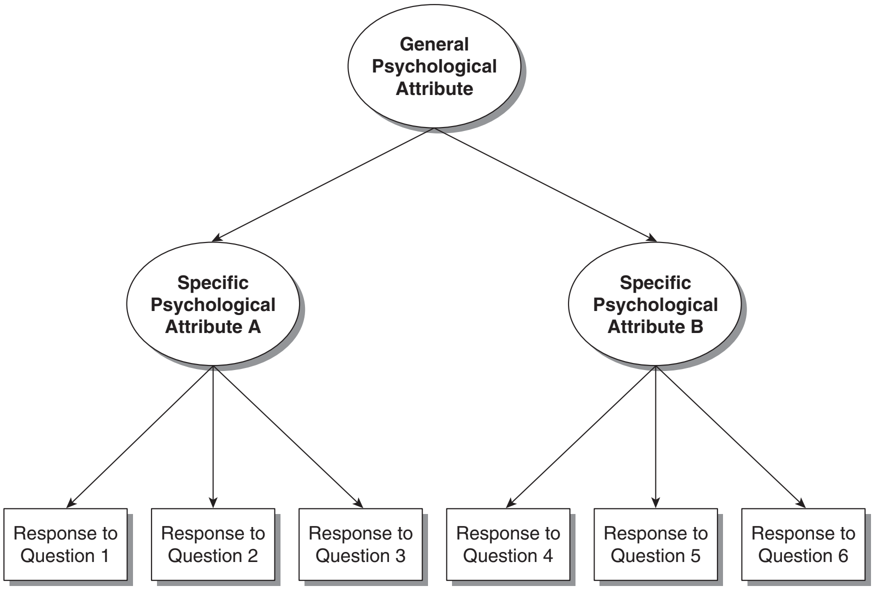


Figure 4.3 Multidimensional Test With Correlated Dimensions (i.e., a Higher-Order Factor Test)

"full-scale" score representing general intelligence, or *g*. We can think of *g* (a general psychological attribute) as affecting a variety of more specific psychological attributes, which in turn affect the way people answer the test questions.

This type of test structure is presented in Figure 4.3. Note that there are two levels of psychological attributes. Responses to each test question are affected by a specific attribute, or factor. For example, an individual's responses to the questions on the quantitative processing subtest of the SB5 are affected by his or her psychological ability to process quantitative information. In contrast, an individual's responses to the questions on the visual–spatial processing subtest of the SB5 are affected by his or her psychological ability to process visual–spatial information. In addition to these specific psychological attributes, there is a general psychological attribute affecting each specific attribute. For example, an individual's abilities to process quantitative information and to process visual–spatial information are partially determined by his or her general cognitive ability, or intelligence. This general attribute is often called a *higher-order factor* because it is at a more general level (or "order") than the specific factors or attributes.

In terms of test evaluation, multidimensional tests are different from unidimensional tests. Recall that a unidimensional test has one and only one score and this score is evaluated with regard to its psychometric quality. In contrast, multidimensional tests have a score for each subtest, and each subtest score is evaluated with regard to its psychometric quality. It is possible that a multidimensional test could have some subtests that have reasonable psychometric quality and other subtests that have poor psychometric quality. Therefore, each subtest requires psychometric examination. For example, the developers and users of the SB5 have examined carefully the reliability and validity of each of its five subtests. In addition, a multidimensional test with correlated dimensions may have a total test score that is computed across its subtests. Thus, this total score also requires psychometric evaluation. For example, the developers and users of the SB5 have examined the reliability and validity of its full-scale score.

In terms of test use, multidimensional tests offer a variety of options. Test users could use any or all of the subtest scores, depending on their relevance to the research or practical context. In addition, test users could use a total test score from a test with correlated dimensions if such a score is computed and has acceptable psychometric properties.

Multidimensional Tests With Uncorrelated Dimensions

As we discussed, the second dimensionality question regards the degree to which a multidimensional test's dimensions are associated with each other (see Figure 4.1). If a test's dimensions are not associated with each other (or are only weakly associated with each other), then the test can be considered a *multidimensional test with uncorrelated dimensions.*

Several personality tests are multidimensional with dimensions that are generally treated as if they are uncorrelated. For example, a test called the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) is a 60-item questionnaire reflecting five dimensions, or factors of personality. That is, the NEO-FFI is designed to measure five relatively independent personality attributes, and these five attributes are not typically treated as reflecting any higher-order factors. Test takers receive five scores—one for each dimension—and each one is itself treated as if it were unidimensional. In a sense, such tests could be viewed as a set of unrelated unidimensional tests that are presented with their items mixed together.

With regard to scoring, evaluation, and use, multidimensional tests with uncorrelated dimensions are similar to multidimensional tests with correlated dimensions, with one important exception. For tests with uncorrelated dimensions, no total test score is computed. That is, a score is obtained for each dimension, but the dimensions' scores are not combined to compute a total test score. Furthermore, each of the dimension scores is evaluated in terms of psychometric quality, and each is potentially used by researchers and practicing psychologists. For example, the NEO-FFI produces only five scores—one for each of the five factors or dimensions; however, no total test score is computed for the NEO-FFI.

This type of test structure is presented in Figure 4.4. Similar to the multidimensional test presented in Figure 4.3, there are two psychological attributes, each one affecting responses to a set of questions. However, in this figure, the two attributes are not linked together in any way. This implies that the attributes are uncorrelated with each other.

The Psychological Meaning of Test Dimensions

After the first two dimensionality issues are addressed (the number of dimensions reflected in a test's items and the association among multiple dimensions), a third important dimensionality issue needs examination. Specifically, test

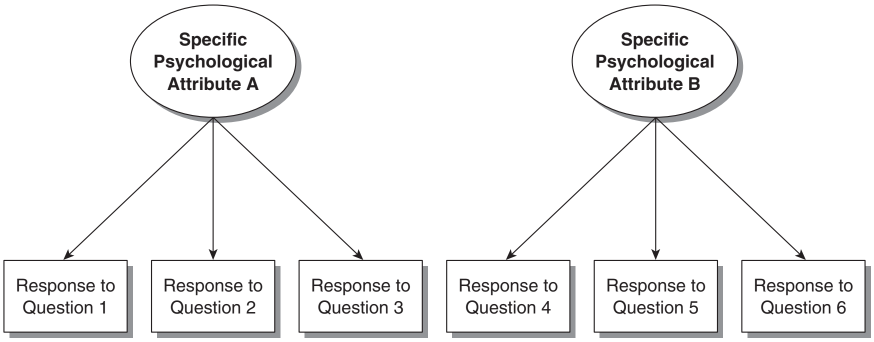


Figure 4.4 Multidimensional Test With Uncorrelated Dimensions

developers, evaluators, and users must understand the psychological meaning of each test dimension. For a test's dimensions to be used and interpreted accurately, test developers and evaluators must conduct research that reveals the psychological attribute that is represented by each test dimension.

In the next section, we discuss a common way in which such research is conducted. We present the basics of a statistical procedure called factor analysis, which is a fundamental tool in the examination of test dimensionality. We present its logic, and we discuss the information that it provides to address each of the three core questions of test dimensionality.

Factor Analysis: Examining the Dimensionality of a Test

Test developers can use a variety of statistical procedures to evaluate a test's dimensionality. Although procedures such as cluster analysis and multidimensional scaling are available, factor analysis is the most common method of examination. By using factor analysis, researchers can address the core questions outlined in the section above, and this provides important insight into the potential scoring, evaluation, and use of psychological tests.

There are, in fact, two broad types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is the more common type, and it is relatively easy to conduct with basic statistical software such as SPSS or SAS. In addition, EFA is often used in early stages of psychometric analysis and development. Considering these issues, the remainder of this chapter focuses primarily on EFA. We will revisit CFA briefly at the end of this chapter, and we will dedicate an entire chapter to it later in this book (Chapter 12).

The Logic and Purpose of Exploratory Factor Analysis: A Conceptual Overview

At the beginning of this chapter, we asked you to consider a six-item personality questionnaire that includes the following adjectives: *talkative*, *assertive*, *imaginative*, *creative*, *outgoing*, and *intellectual*. Furthermore, we asked you to consider the number of different attributes that are reflected in these adjectives. As we mentioned, reasonable people might disagree about this question. Based on one's particular interpretation of the adjectives and one's understanding of personality, one might argue that the six adjectives reflect one single dimension, two dimensions, or perhaps three or more dimensions.

An important difficulty with this approach—an approach based only on our interpretations of the meaning of items—is that it is not easy to evaluate which perspective is the best. That is, if you believe that there is a two-factor structure to the questionnaire but your colleague believes that there is a three-factor structure, then how could you determine who is correct or if either one of you is correct?

Rather than relying on idiosyncratic interpretations of the meaning of items, test developers and users often prefer to base their arguments on empirical data. Therefore, we might give the six-item questionnaire to a sample of 100 respondents, asking each respondent to rate each item in terms of the following response options (circling the number for the appropriate option):

1	*2*	*3*	*4*	*5*
Completely unlike me	Somewhat unlike me	Neither like me nor unlike me	Somewhat like me	Completely like me

We then enter their data into a statistical software computer program and compute the correlations among the six items. We would then use the correlations to help us identify and interpret the dimensions reflected in the items.

For example, take a moment to examine the hypothetical correlation matrix presented in Table 4.1. Note that three of the items—"talkative," "assertive," and "outgoing"—are all strongly correlated with each other. An individual who rates herself as relatively high on one of these three items is likely to rate herself as relatively high on the other two items. We also see that the other three items—"imaginative," "creative," and "intellectual"—are strongly correlated with each other. Importantly, we also see that these two clusters of items are independent. For example, the correlation between "talkative" and "creative" is 0, as is the correlation between talkative and imaginative, between outgoing and intellectual, and so on. That is, the fact that an individual rates himself as assertive, talkative, and outgoing says nothing about that person's likely level of creativity, imagination, or intellect. This pattern of correlations begins to reveal the dimensionality of the six-item personality test.

By scanning an interitem correlation matrix in this way, we could begin to understand a test's dimensionality. Essentially, we try to identify sets of items that go together—sets of items that are relatively strongly correlated with each other but weakly correlated with other items. Each set of relatively highly correlated items represents a psychological dimension or "factor."

Indeed, we can begin to address the three dimensionality questions in Figure 4.1. To determine the number of factors within a scale, we count the number of sets that we identify. If all scale items are well correlated with each other at approximately equal levels, then there is only a single set (i.e., factor) and the scale is unidimensional. If, however, there are two or more sets, then the scale is multidimensional. We identified two sets of items in the hypothetical correlation matrix in Table 4.1—these findings suggest that the six-item personality questionnaire has a two-dimensional structure (i.e., it is multidimensional). That is, three items cluster together into one dimension, and the other three cluster into a second dimension.

To determine whether the factors are correlated with each other, we would examine the pattern of correlations between the sets. That is, the potential correlations between factors are based on the correlations between items in different sets. In the example in Table 4.1, we found that the items from one set were uncorrelated with the items in the other set. This suggests that the two factors are, themselves, uncorrelated with each other—the factor represented by the items in one set is unrelated to the factor represented by the items in the other set. Thus, this six-item test appears to be a multidimensional test with uncorrelated dimensions. However, if items from one set are, in fact, correlated with items from another set, then the factors are correlated with each other. For example, we might have found a correlation of .30 between talkative and creative, a correlation of .25 between talkative and imaginative, a correlation of .32 between outgoing and intellectual, and so on. Such a pattern of moderately sized cross-factor correlations would suggest that the factors are correlated with each other—that we were working with a test that was multidimensional with correlated dimensions.

Finally, to understand the potential psychological meaning of the factor, we examine the content of the items constituting that factor. That is, a factor's potential meaning arises, in part, from the psychological concept or theme that its items share. Consider, for example, the items "talkative," "assertive," and "outgoing." What do they have in common? What is a common psychological concept that they share? Many personality psychologists would likely suggest that these items reflect an extraversion factor and that the other three items (i.e., "imaginative," "creative," and "intellectual") reflect openness to experience. The answer to this question is, of course, based on interpretation, judgment, and preference. Indeed, one person's answer might differ from another's. The interpretation of "extraversion" and "openness to experience" is based on familiarity with "the five-factor model" of personality, which is widely known in personality psychology and which includes the traits of extraversion and openness to experience. People with other perspectives and backgrounds might choose to label the factors differently.

By examining the pattern of correlations in this way, we have performed a very basic factor analysis. Unfortunately, such a simplistic "eyeballing" approach rarely works with real data. Real data usually include many more items. In the current example, we examined only six items, but many measures include considerably more than six items. For example, the Conscientiousness scale of the NEO Personality Inventory–Revised (NEO-PI–R) questionnaire (Costa & McCrae, 1992) includes 48 items. Difficulty arises because a larger number of items produces a much larger number of correlations to examine. For example, if we examined a correlation matrix for 48 items, we would have to inspect more than 1,100 correlations! Obviously, visually inspecting such a large correlation matrix is a nearly impossible task. In addition to the large number of correlations in most real data, the pattern of correlations in real data is never as clear as it appears to be

Table 4.1 (Hypothetical) Correlation Matrix for a Two-Factor Set of Items

	Talkative	*Assertive*	*Outgoing*	*Creative*	*Imaginative*	*Intellectual*
Talkative	1.00					
Assertive	.66	1.00				
Outgoing	.54	.59	1.00			
Creative	.00	.00	.00	1.00		
Imaginative	.00	.00	.00	.46	1.00	
Intellectual	.00	.00	.00	.57	.72	1.00

in Table 4.1. The hypothetical correlations in Table 4.1 include a few very strong positive correlations and a few zero correlations, but nothing else. In real data, correlations often are closer to .18 or –.32 than to .70. Therefore, the clusters of items in real data are much more ambiguous than the ones in Table 4.1, and this ambiguity complicates the process of evaluating dimensionality.

EFA is a statistical procedure that simplifies this process. Rather than visually inspecting a matrix of dozens or even hundreds of correlations, we can use EFA to process a large set of correlations. Because the analysis of dimensionality typically is "messier" than the example in Table 4.1, researchers often rely on EFA to examine the dimensionality of psychological measures.

Conducting and Interpreting an Exploratory Factor Analysis

Factor analysis can be conducted by using participants' raw data—their responses to each individual item in a test. However, some statistical software packages allow factor analysis to be conducted on a correlation matrix that summarizes the associations among test items. Thus, if you have access to the appropriate software, you could replicate the example analyses that we report and interpret below.

Figure 4.5 is a flowchart of the process of conducting an EFA. As this figure illustrates, EFA is often an iterative process, as the results of one step often lead researchers to reevaluate previous steps.

Choosing an Extraction Method. In the first step of an EFA, we choose an "extraction method." This refers to the specific statistical technique to be implemented, and options include principal axis factoring (PAF), maximum likelihood factor analysis, and principal components analysis (PCA), among others.

PAF and PCA are the common choices in most applications of EFA. Although PCA is not technically a "factor" analysis, it is essentially the same thing and is the default method for several popular statistical software packages' factor analysis procedure. Although the results obtained from PAF are often quite similar to those obtained from PCA, many experts recommend PAF over PCA. For example, Fabrigar, Wegener, MacCallum, and Strahan (1999) conclude that PCA is not recommended "when the goal of the analysis is to identify latent constructs underlying measured variables" (p. 276), as is typically the case in psychometric evaluation.

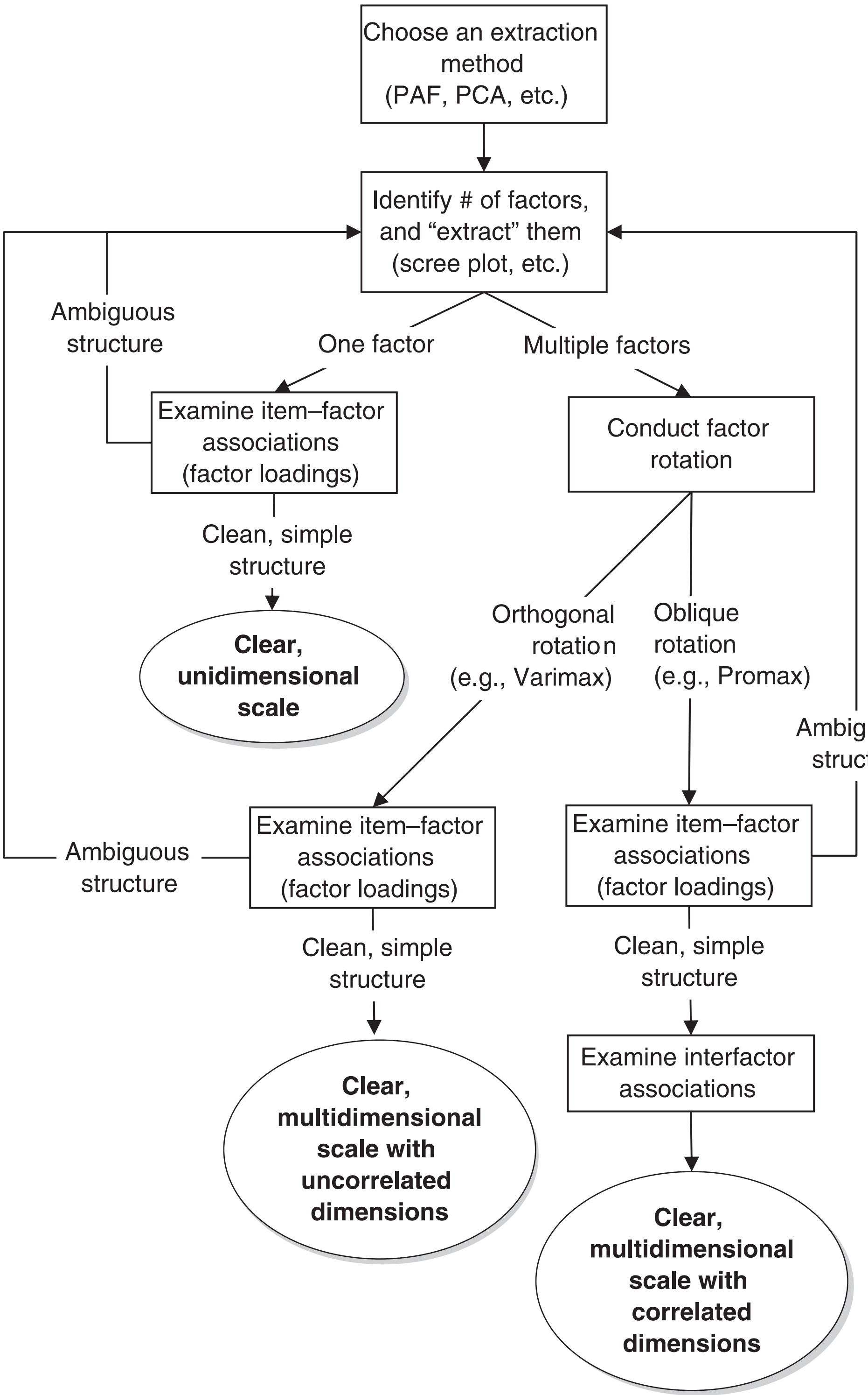


Figure 4.5 Process Flowchart of an Exploratory Factor Analysis

NOTE: PAF = principal axis factoring; PCA = principal components analysis.

To illustrate the EFA process, we will use a PAF extraction method to analyze the data illustrated in Table 4.1. Note that responses should be reverse scored, if necessary, before conducting the EFA (see Chapter 10).

Identifying the Number of Factors and Extracting Them. In the second step of an EFA, we identify the number of factors within our set of items, and we direct the statistical software to "extract" that number of factors. Unfortunately, we have no single, simple rule that we can use to make this identification. Instead, we must rely on rough guidelines and subjective judgment.

To address the "number of factors" issue, test developers and test users often refer to statistics called *eigenvalues.* In Figure 4.6, this information is presented in the "Total Variance Explained" box—specifically, the six eigenvalues are in the "Total" column under the "Initial Eigenvalues" heading. Although there are highly technical definitions of eigenvalues, what matters for our current discussion is how eigenvalues are used, not what they are. There are many ways in which this information can be used (e.g., parallel analysis; see Hayton, Allen, & Scarpello, 2004), but we will focus on the three ways that are the most common and that are integrated into most popular statistical software options.

One way of using eigenvalues is to examine the relative sizes of the eigenvalues themselves. Note that the eigenvalue output in Figure 4.6 includes six rows. Each row represents the potential number of dimensions reflected among the six test items. That is, this output will always include a number of rows that is equal to the number of items on the test, and each item might reflect a different dimension.

Examining the eigenvalues, we scan down the descending values in this column, and we hope to find a point at which all subsequent differences between values become relatively small. For example, in our output, we see a relatively large difference between the second eigenvalue (2.173) and the third eigenvalue (0.563). We also note that this difference is much larger than all the subsequent other row-by-row differences. That is, the difference between the third and fourth eigenvalues is small, as is the difference between the fourth and fifth, and so on.

The "location" of this point has implications for the answer to the "number of dimensions" question. We find this point, and we conclude that the test has a number of dimensions equal to the row with the larger eigenvalue. In Figure 4.6, the point is located between Rows 2 and 3, so we would conclude that the test has two dimensions. If the large difference was located between Rows 1 and 2, then we would conclude that the test has one dimension (i.e., that the test is unidimensional). Similarly, if the large difference was located between Rows 4 and 5, then we would conclude that the test has four dimensions.

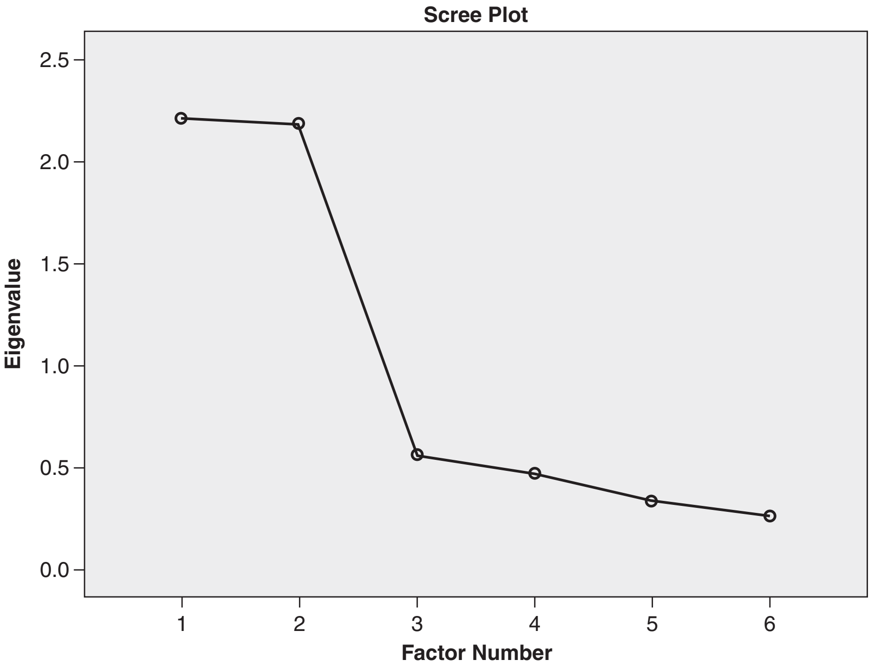
Although it has been criticized, the "eigenvalue greater than 1.0" rule is the second common way in which eigenvalues are used to evaluate the number of dimensions. As represented by the fact that several popular statistical packages (e.g., SPSS and SAS) use this as a default option for answering the "number of dimensions" question, many factor analysts base their judgments on the number of eigenvalues that are greater than 1.0. For example, of the six eigenvalues in Figure 4.6, only two are above 1.0. Therefore, we might conclude that the test items reflect two dimensions. If our analysis had revealed three eigenvalues greater than 1.0, then we might conclude that the test items reflect three dimensions.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.195	36.578	36.578	1.836	30.599	30.599	1.836
2	2.173	36.222	72.800	1.808	30.131	60.730	1.808
3	.563	9.382	82.183				
4	.472	7.867	90.050				
5	.333	5.554	95.604				
6	.264	4.396	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.



(Continued)

<AQ: is Factor Matrix in 2nd column correct? Or should it be Pattern Matrix?>

(Continued)

Factor Matrix[a]

	Factor	
	1	2
Intellectual	.942	.000
Imaginative	.764	.000
Creative	.604	.000
Assertive	.000	.849
Talkative	.000	.777
Outgoing	.000	.695

Extraction Method: Principal Axis Factoring.

[a] 2 factors extracted. 19 iterations required.

Factor Matrix[b]

	Factor	
	1	2
Intellectual	.942	.000
Imaginative	.764	.000
Creative	.604	.000
Assertive	.000	.849
Talkative	.000	.777
Outgoing	.000	.695

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

[b] Rotation converged in 2 iterations.

Structure Matrix

	Factor	
	1	2
Intellectual	.942	.000
Imaginative	.764	.000
Creative	.604	.000
Assertive	.000	.849
Talkative	.000	.777
Outgoing	.000	.695

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix

Factor	1	2
1	1.000	.000
2	.000	1.000

Extraction Method: Principal Axis Factoring.

Rotation Method: Promax with Kaiser Normalization.

Figure 4.6 Selected Output From Exploratory Factor Analysis of the Correlations in Table 4.1

Again, we should note that, despite its popularity, the "eigenvalue greater than 1.0" rule has been criticized as inappropriate for evaluating the number of dimensions in many applications of factor analysis (Fabrigar et al., 1999). Indeed, this guideline "is *among the least accurate methods* for selecting the number of factors to retain" (Costello & Osborne, p. 2), and it should generally *not* be used as a guideline for identifying the number of factors.

A third common way of using eigenvalues is to examine a *scree plot*, and it is probably the best of the three most common methods of identifying the number of factors. As illustrated by Figure 4.6's presentation of the scree plot resulting from our EFA, a scree plot is a graphical presentation of eigenvalues. Similar to the examination of eigenvalues discussed above, we look for a relatively large difference or drop in the plotted values. More specifically, we hope to find an obvious "leveling-off point" in the plot (as we move from left to right along the *x*-axis).

For example, the scree plot in Figure 4.6 shows an obvious flattening beginning at Factor 3. An obvious flattening point suggests that the number of factors is one less than the factor number of the flattening point. That is, if there is a flattening point beginning at the second eigenvalue, then this indicates the presence of only one factor. In contrast, if a flattening point begins at the third eigenvalue (as in our scree plot), then this indicates the presence of two factors, and so on.

If we do obtain a clear answer to the "number of factors" question, then we extract that number of factors. In most software programs, this simply means that we tell the program to proceed to the next step (see the flowchart in Figure 4.5) using the number of factors that we have identified. In the case of the data in Table 4.1 and Figure 4.6, we directed the program to proceed with two factors. We will return to the EFA of these data shortly.

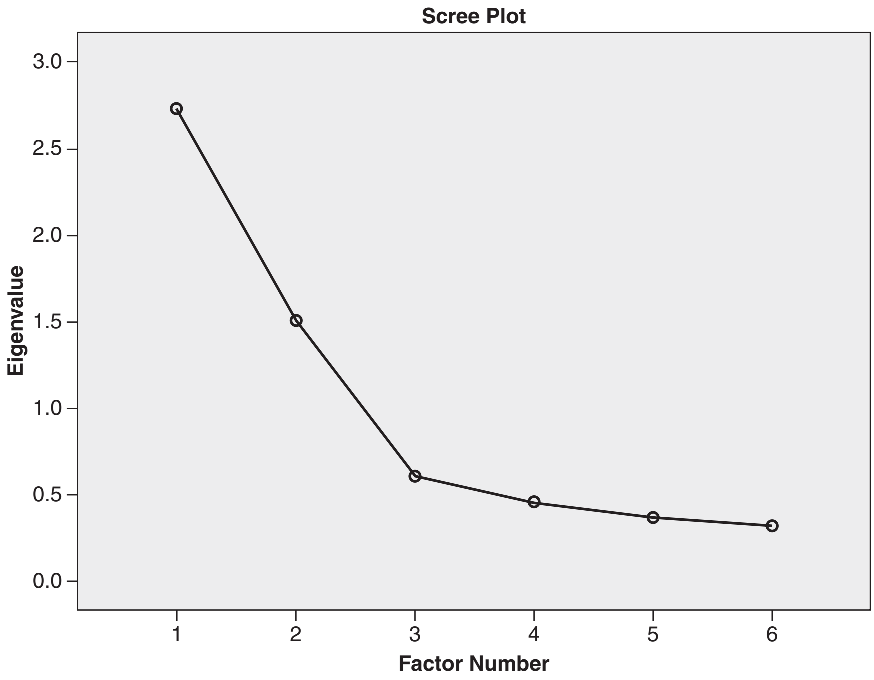
Unfortunately, scree plots are not always clear—certainly rarely as clear as the one in Figure 4.6, which was based on hypothetical data constructed to be as obvious as possible. Providing more realistic examples, Figures 4.7 and 4.8 show selected results of two additional EFAs based on hypothetical data from two different six-item scales. Figure 4.7 is more realistic than the results in Figure 4.6, but it is still fairly clear—we see a relatively clear flattening point at the third eigenvalue, again indicating a two-dimensional structure to the items. In contrast, the scree plot in Figure 4.8 is extremely ambiguous—there is no clear flattening point that would guide our decision about the number of factors. In ambiguous cases like this, we use additional information to guide our understanding of the scale's number of dimensions.

One type of additional information is the clarity with which the scale's items are associated with its factors. For example, the ambiguous scree plot in Figure 4.8 might lead us to (somewhat arbitrarily) extract two factors and examine the item–factor associations in the next step of the EFA. As we will discuss shortly, our results from that later step might motivate us to revisit the present step, extract a different number of factors, and proceed again to the next steps. This is the iterative back-and-forth nature of EFA that was mentioned earlier.

Occasionally, we never obtain a clear answer to the "How many dimensions?" question, suggesting that the scale has no clear dimensionality. If we did encounter that situation, then we might conclude that the scale needs revision—for example, in terms of clarifying the construct(s) that it is intended to assess or in terms of revising the items themselves (see Furr, 2011, chaps. 2 and 3).

In a typical EFA, researchers make an initial decision about the scale's number of factors and then move on to one of two subsequent steps. As illustrated in the flowchart in Figure 4.5, if the scree plot (or another good guideline) suggests a single dimension, then researchers proceed directly to examining the associations between the items and that factor. We will discuss this later. However, if there is evidence of more than one dimension, then researchers next make decisions about rotating the factors.

Rotating the Factors. If the evidence suggests that a scale is multidimensional, then we usually "rotate" the factors. The purpose of this step is to clarify the psychological meaning of the factors.



Factor Matrix[a]

	Factor	
	1	2
item6	.754	−.457
item2	.667	.536
item5	.620	−.399
item4	.589	−.339
item1	.561	.417
item3	.513	.391

Extraction Method: Principal Axis Factoring.

[a] 2 factors extracted. 15 iterations required.

Pattern Matrix[a]

	Factor	
	1	2
item6	.880	.004
item5	.743	−.017
item4	.672	−.021
item2	−.015	.861
item1	.016	.694
item3	.007	.642

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

[a] Rotation converged in 3 iterations.

Structure Matrix

	Factor	
	1	2
item6	.882	.314
item5	.737	.245
item4	.679	.257
item2	.288	.855
item1	.260	.699
item3	.233	.644

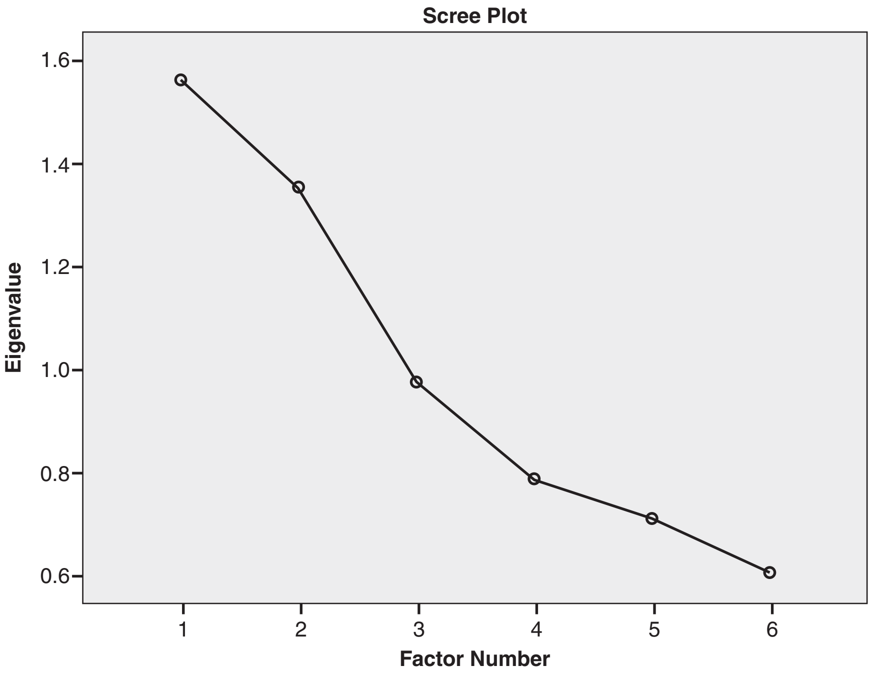
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix

Factor	1	2
1	1.000	.352
2	.352	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Figure 4.7 Selected Output From Exploratory Factor Analysis of "More Realistic" Data From a Six-Item Questionnaire



Factor Matrix[a]

	Factor	
	1	2
item5	.959	−.138
item4	.359	.022
item1	.069	−.564
item3	.098	−.471
item2	.032	.368
item6	.164	.266

Extraction Method: Principal Axis Factoring.

[a] 2 factors extracted. 206 iterations required.

Pattern Matrix[a]

	Factor	
	1	2
item5	.969	−.004
item4	.358	.028
item1	−.024	.568
item3	.020	.480
item2	−.029	.368
item6	.118	.286

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

[a] Rotation converged in 3 iterations.

Structure Matrix

	Factor	
	1	2
item5	.969	.021
item4	.359	.037
item1	−.010	.568
item3	.032	.481
item2	−.020	.368
item6	.126	.289

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization.

Factor Correlation Matrix

Factor	1	2
1	1.000	.026
2	.026	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

Figure 4.8 Selected Output From Exploratory Factor Analysis of Ambiguously Structured Data From a Six-Item Questionnaire

There are two general types of rotation, and they have differing implications for the potential associations among factors. The first general type of rotation is an orthogonal rotation, and it generates factors that are uncorrelated or "orthogonal" to each other. A procedure called "varimax" is the standard orthogonal rotation. The second general type of rotation is an oblique rotation, which generates factors that can be either correlated or uncorrelated with each other. There are many subtypes of oblique rotations, including "promax" and "direct oblimin." A full discussion of the differences among these subtypes is beyond the scope of our discussion—the important point is that all the oblique rotations allow factors to be correlated or uncorrelated with each other. To anthropomorphize, if factors "want to be" correlated with each other, then oblique rotations allow them to be correlated; and if factors "want to be" uncorrelated, then oblique rotations allow them to be uncorrelated.

Many experts suggest that oblique rotations are preferable to orthogonal rotations (e.g., Fabrigar et al., 1999). Again, the main purpose of rotation is to clarify the nature of the factors, which (as we will discuss next) depends on the pattern of associations between the factors, on one hand, and the scale's items, on the other. Oblique rotations can produce results in which these associations are as clear as possible, allowing us to understand our scales as clearly as possible. With this in mind, there is often little conceptual or psychometric reason to force a scale's factors to be orthogonal (i.e., uncorrelated)—doing so can create less clarity about the scale as compared with oblique rotations. After rotating factors, we next examine the associations between the items and the factors.

Examining Item–Factor Associations. Although a full understanding of a scale's dimensions emerges from many kinds of information (as discussed in later chapters on reliability and validity), the associations between items and factors can be an important piece of the puzzle. EFA presents these associations in terms of "factor loadings," and each item has a loading on each factor. By examining the loadings and identifying the items that are most strongly linked to each factor, we can begin to understand the factors' psychological meaning.

Generally, factor loadings range between –1 and +1, and they are interpreted as correlations or as standardized regression weights. When using an orthogonal rotation (or when a scale has only one factor), we obtain loadings that can be seen as correlations between each item and each factor. In contrast, when using oblique rotations, we obtain several kinds of factor loadings. For example, if we use the statistical program SPSS and we choose an oblique rotation, then we obtain both "pattern coefficients" and "structure coefficients." Pattern coefficients reflect the "unique association" between an item and a factor. That is, a pattern coefficient reflects the degree to which an item is associated with a factor, controlling for the correlation between the factors. For readers who are familiar with multiple regression, pattern coefficients are the standardized regression weights produced by a regression analysis in which respondents' item responses are predicted from their levels of the underlying factors. In contrast, structure coefficients are simply correlations between respondents' item responses and their levels of the underlying factors. By controlling for any correlation between factors, pattern coefficients can provide sharper clarity about the unique associations between items and factors as compared with structure coefficients.

When interpreting factor loadings, two pieces of information are important (see our discussion of interpreting correlations and covariances in Chapter 3). First, the *size* of the loading indicates the degree of association between an item and a factor—larger loadings (i.e., loadings farther from 0, closer to –1 or +1) indicate stronger associations between an item and a factor. More specifically, loadings above .30 or .40 are often seen as reasonably strong, with loadings of .70 or .80 being seen as very strong. The second important piece of information is the direction of a loading—positive or negative. A positive loading indicates that people who respond with a "high score" on the item have a high level of the underlying factor. In contrast, a negative loading indicates that people who respond with a high score on the item have a low level of the underlying factor.

For example, recall that the scree plot in Figure 4.6 strongly indicated the presence of two factors. With this in mind, we continued our EFA of these data by extracting two factors and using an oblique rotation (i.e., "Promax"). We obtained the loadings also shown in 4.6; in fact, there are three sets of loadings. The "Factor Matrix" presents the factor loadings that would be obtained before rotating the factors. Given the usefulness of factor rotations, we generally ignore these loadings. As the "Pattern Matrix" label implies, the second set of loadings is the pattern coefficients. And, of course, the "Structure Matrix" presents the structure coefficients.

Examining all three matrices reveals a very clear pattern of item–factor associations. Indeed, these results are highly consistent with our earlier "eyeball" factor analysis of the correlations in Table 4.1. Specifically, the items "intellectual," "imaginative," and "creative" load positively and strongly on Factor 1—the lowest loading being .604. Similarly, the items "assertive," "talkative," and "outgoing" load strongly and positively on Factor 2. Importantly, the first set of items (i.e., "intellectual," etc.) do not load on Factor 2, and the second set of items do not load on Factor 1.

Note that the three sets of loadings in Figure 4.6 are identical. That is, the Factor Matrix, Pattern Matrix, and Structure Matrix have identical values. This is a very atypical finding that, again, results from the fact that the correlations in Table 4.1 were created to be as clear and simple as possible. Thus, these results are rather artificial—in real analyses of oblique rotations, these matrices will differ from each other. We will illustrate this shortly.

The factor loadings in Figure 4.6 are an ideal example of "simple structure." Simple structure occurs when each item is strongly linked to one and only one factor. Again, in Figure 4.6, each item loads robustly on one factor but has a loading of .000 on the other factor. Thus, each item clearly belongs on one and only one factor.

Simple structure is important in psychometrics and scale usage. Generally, we sum or average a respondent's responses to the items that load together on a factor. For example, if we used the six-item questionnaire analyzed in Figure 4.6, we would create two scores for each person. Recall that for our six-item questionnaire, we asked each hypothetical respondent to rate himself or herself on each item, using a 5-point set of response options (i.e., 1 = *Completely unlike me*, 5 = *Completely like me*). First, we would sum (or average) a person's responses to intellectual, imaginative, and creative, producing an "openness to experience" score for each person (based only on those three items). Second, we would combine each person's responses to "assertive," "talkative," and "outgoing," producing an "extraversion" score for each person. Note that if an item does not load on a factor, then it is not included in scoring of that dimension/factor. Thus, simple structure is important because it reveals which items should be scored together. Because rotation makes it more likely that we will obtain simple structure, rotation is usually a key part of EFA.

For a more realistic example, consider the EFA results in Figure 4.7, conducted on a different (hypothetical) set of six items. As noted earlier, this scree plot indicates two factors; thus, we extracted two factors, used an oblique rotation, and obtained factor loadings. There are several important points to note. First, note that the three matrices differ from each other. Again, as noted earlier, this is typical—the Factor Matrix will differ from the Pattern Matrix, which will differ from the Structure Matrix. Second, note that the loadings in the Factor Matrix do not show simple structure. That is, each item loads fairly robustly on *both* factors. In this case, the lack of simple structure occurs because the Factor Matrix includes factor loadings that are obtained before rotation has taken place. Thus, as mentioned earlier, we typically ignore these results, even though they are often provided by the statistical package. Third, the Pattern Matrix does show very good simple structure—each item loads robustly on one and only one item. Fourth, the loadings in the Structure Matrix have a somewhat less clear simple structure than the loadings in the Pattern Matrix. This result is pretty typical, and it arises from the difference (discussed earlier) between pattern coefficients and structure coefficients. Fifth, we now see negative factor loadings, although none of the negative loadings in the Pattern Matrix are large enough to be very meaningful. As compared with the highly artificial results in Figure 4.6, the results in Figure 4.7 are a much more realistic illustration of a clear two-factor scale with very good simple structure.

For a full understanding of item–factor associations, it is important to realize that factor loadings can violate simple structure in two ways. First, an item might not load strongly on any factor, and second, an item might load strongly on more than one factor.

For example, consider again the results in Figure 4.8, illustrating a dimensionality that appears quite unclear. Based on another hypothetical six-item questionnaire, the scree plot is highly ambiguous (as discussed earlier). Because of this ambiguity, we rather arbitrarily tried a two-factor extraction, and we used an oblique rotation. Concentrating on the Pattern Matrix, we see one clear problem—Item 6 does not load very strongly on either factor (i.e., loadings below .30 on both). Two other slight problems are that the strongest factor loadings for Items 4 and 2 are below .40—ideally, an item would have an even stronger factor loading. Such results create ambiguity with regard to this scale. Do all of these items belong on the questionnaire? How should the questionnaire be scored? Are there really two factors, perhaps more, perhaps less?

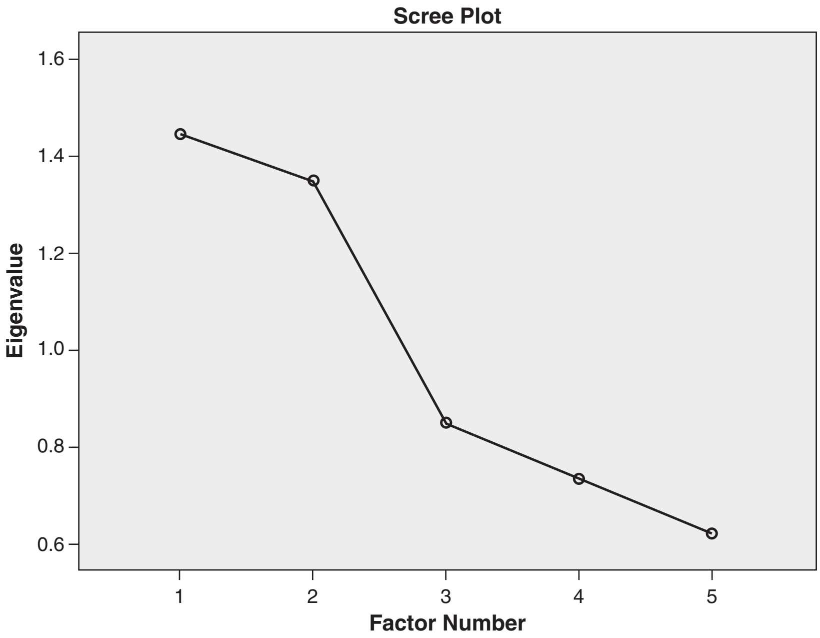
As shown in the EFA flowchart (Figure 4.5), when faced with such ambiguity, one option is to revisit our initial decision about the number of factors to extract. We noted earlier that scree plots sometimes fail to provide clear information about this issue, but the item–factor associations might help shape our decision about the number of factors. Revisiting again the unclear structure in Figure 4.8, we examined factor loadings based on several different factor structures. Our hope was to find a number of factors that produces factor loadings with a clear simple structure. If we find loadings that are relatively clear and meaningful, then we might decide that the "correct" number of factors is the one producing that pattern of factor loadings. In our analysis of the ambiguous data represented in Figure 4.8, we also examined one-factor and three-factor extractions. Unfortunately, neither analysis produced clearer results.

Failing to find a clearer solution by revisiting the number of factors, there is at least one additional option for dealing with factorial ambiguity. Specifically, we might drop items that have poor structure. If an item is not strongly associated with any factor, then we conclude that it simply is not coherently related to the other items on the test or questionnaire. This might suggest that the item reflects a psychological construct that differs from the one(s) reflected by the other items on the scale (e.g., having a single math item on a vocabulary test). Alternatively, it might suggest that the item is strongly affected by random measurement error (see the later chapters on reliability). Either way, the item, as it is, likely does not belong on the scale. We noted that another problem is when an item loads robustly on more than one factor. In such cases, the item reflects more than one psychological construct. That is, responses to the item are affected by several psychological traits, abilities, or states (or what have you). Such an item does not uniquely reflect any construct, and thus we might drop it or revise it to reflect only one construct.

With this option in mind, we revisited the data reflected in Figure 4.8's ambiguous results. Noting that Item 6 seemed to load weakly on both factors (in the two-factor solution), we removed this item from the analysis and reconducted the EFA. Essentially, this addresses the dimensionality of a questionnaire that would include only Items 1 through 5. Figure 4.9 presents the results of this analysis, showing that this adjustment produces a questionnaire that now has a clearer dimensionality. Indeed, the scree plot now clearly suggests two factors, and the factor loadings have good simple structure—each of the five remaining items loads on one and only one factor. Apparently the inclusion of Item 6 created ambiguity in the questionnaire as a whole. Thus, by dropping that item from the questionnaire, we are left with a five-item questionnaire that clearly includes two dimensions.

Examining the Associations Among Factors. Finally, as shown in the EFA flowchart (Figure 4.5) when using oblique rotations, we should examine the correlations among the factors. Recall that oblique rotations allow factors to be either correlated or uncorrelated with each other, whereas orthogonal rotations force the factors to be uncorrelated. The results of oblique rotations thus include a correlation for each pair of factors, revealing the higher-order associations among factors. This information has implications for our understanding of the nature of the factors and for the scoring of the test or questionnaire. As mentioned earlier (Figure 4.1), we should create "total scores" from a multidimensional scale only when the dimensions are correlated with each other to a meaningful degree.

Returning to our first and main example (see Table 4.1 and Figure 4.6), the factor correlation is presented in the "Factor Correlation Matrix" box. This small matrix presents the correlation between the two factors that we extracted and



Factor Matrix[a]

	Factor 1	Factor 2
item4	.846	–.117
item5	.401	–.082
item1	.137	.651
item2	.030	.394
item3	.084	.368

Extraction Method:
Principal Axis Factoring.
[a] 2 factors extracted.
279 iterations required.

Pattern Matrix[a]

	Factor 1	Factor 2
item4	.854	.006
item5	.410	–.024
item1	.009	.665
item2	–.046	.395
item3	.011	.376

Extraction Method:
Principal Axis Factoring.
Rotation Method:
Promax with Kaiser
Normalization.
[a] Rotation converged in
3 iterations.

Structure Matrix

	Factor 1	Factor 2
item4	.854	.048
item5	.409	–.003
item1	.042	.665
item2	–.026	–.362
item3	.030	.377

Extraction Method:
Principal Axis Factoring.
Rotation Method:
Promax with Kaiser
Normalization.

Factor Correlation Matrix

Factor	1	2
1	1.000	.050
2	.050	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.

ED: indent turnovers under charts?

Figure 4.9 Selected Output From Exploratory Factor Analysis of Data From a Five-Item Version of the Questionnaire Originally Analyzed in Figure 4.8

rotated earlier in the analysis. This output reveals a zero correlation between the two dimensions, indicating that the two dimensions are not associated with each other. That is, people who have a high level of openness to experience are not particularly likely (or particularly unlikely) to have a high level of extraversion.

Again, it is important to note that different data will produce different results—it is quite possible that an oblique rotation will produce dimensions that are more highly correlated with each other. For example, Figure 4.7 presents a two-factor structure in which the two factors are indeed more highly correlated, at .35. This suggests that people who have a relatively high level of the first psychological dimension are likely to have a relatively high level of the second dimension.

In sum, oblique rotations allow factors to be correlated "however they want to be." For the questionnaire represented in Figure 4.6, the factors "wanted" to be uncorrelated, and the oblique rotation allowed them to be uncorrelated (i.e., the interfactor correlation was .00). In contrast, for the questionnaire represented in Figure 4.7, the factors "wanted" to be correlated, and the oblique rotation allowed them to be correlated.

For some final insights into the links between rotations, factor correlations, and factor loadings, consider what happens if we use an orthogonal rotation for these EFAs. In the case of the data from Figure 4.6 (the original, highly artificial data), varimax rotation produces the factor loadings shown in Figure 4.10a. Note that these loadings are identical to those obtained in the original analysis based on an oblique rotation (see Figure 4.6). In the case of the data from Figure 4.7 (the more realistic data in which the factors were moderately correlated with each other), varimax rotation produces the loadings given in Figure 4.10b. Note that these loadings differ from those obtained in the analysis based on an oblique rotation (see Figure 4.7).

Take a moment to consider why this might be—why in one case orthogonal and oblique rotations produce the same result, whereas in the other case they produce different results. The answer is that for the data in Figure 4.6, the factors "want" to be uncorrelated. That is, the oblique rotation (Figure 4.6) allowed the factors to be either correlated or uncorrelated, and the results showed that the factors were "naturally" uncorrelated. Because the oblique rotation produced results in which the factors were uncorrelated, the orthogonal rotation (which forces the factors to be uncorrelated) produced the exact same results in terms of factor loadings. In contrast, the questionnaire reflected in Figure 4.7 includes factors that "want" to be correlated—the oblique rotation allowed them to be correlated, and they were, in fact, correlated at .35. When we then conducted an orthogonal rotation, which forced the factors to be uncorrelated, this changed the nature of the factors, which then changed the associations between the items and the factors (i.e., it changed the factor loadings).

Also, notice the way the factor loadings changed—compare the pattern coefficients in Figure 4.7 with the factor loadings in Figure 4.10b. This comparison reveals that the orthogonal rotation produced factor loadings that are somewhat less clear—they have worse simple structure. For example, examine the loadings for Item 6 (the top item in the matrix). In the oblique rotation, its loading on Factor 1 was .880, and its loading on Factor 2 was .004. In the orthogonal rotation, its

(a)

Rotated Factor Matrix[a]

	Factor	
	1	2
Intellectual	.942	.000
Imaginative	.764	.000
Creative	.604	.000
Assertive	.000	.849
Talkative	.000	.777
Outgoing	.000	.695

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization.

[a] Rotation converged in 2 iterations.

(b)

Rotated Factor Matrix[a]

	Factor	
	1	2
item6	.867	.162
item5	.728	.117
item4	.665	.141
item2	.139	.844
item1	.140	.685
item3	.121	.633

Extraction Method: Principal Axis Factoring. Rotation Method: Varimax with Kaiser Normalization.

[a] Rotation converged in 2 iterations.

Figure 4.10 Factor Loadings From Orthogonal Rotation of Data From (a) Figure 4.6 and (b) Figure 4.7

loading on Factor 1 is weaker, at .867 (though still quite strong), and its loading on Factor 2 is somewhat stronger, at .162 (though still relatively weak). All the items show this pattern—in the orthogonal rotation, their main or "on-factor" loadings are somewhat weaker, and their other or "off-factor" loadings are somewhat stronger. Thus, orthogonal rotation can produce a somewhat less simple structure within the factor loadings.

A Quick Look at Confirmatory Factor Analysis

As noted earlier, there are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Our discussion so far has focused on EFA because it has been used more frequently than CFA, because it is relatively easy to conduct with basic statistical software, and because it is often used in early phases of the development and evaluation of psychological tests. However, a brief discussion of CFA and how it differs from EFA is potentially useful at this point.

Although both EFA and CFA are approaches to factor analysis, they have somewhat different purposes or roles. As its label implies, EFA is as an exploratory procedure—it is designed for situations in which there are few, if any, ideas about a test's dimensionality. Again, test developers and evaluators might use EFA in the early phases, while conducting basic exploratory analyses of a set of items.

In contrast, CFA is a confirmatory procedure—it is designed for situations in which there are very clear ideas about a test's dimensionality. For example, we might wish to evaluate the dimensionality of a 16-item test that has been developed specifically to have 8 questions reflecting one factor and another 8 questions reflecting a different factor. In this case, we would have a fairly clear idea about the intended dimensionality of the test. That is, we would know exactly which items were intended to load on which factor. After collecting a large number of responses to this 16-item test, we could use CFA to directly test these ideas; or perhaps more accurately, we could test whether the responses to the test items match or fit with these ideas (i.e., whether the test shows the dimensionality that it is intended to have). In this way, CFA is used to confirm, or potentially disconfirm, our hypotheses about a test's dimensionality.

There are important similarities between CFA and EFA, but the process of conducting a CFA is substantially different from the process of conducting an EFA. Indeed, CFA includes new concepts and statistics, such as inferential tests of parameter estimates and "goodness-of-fit" indices. Moreover, although most of the common statistical software packages can now be used to conduct a CFA, the way this is done differs rather dramatically from the way those packages conduct an EFA.

Given the differences between EFA and CFA, and given the additional complexity of CFA, we will return to CFA later in the book, in Chapter 12. In that chapter, we will describe the information provided by a CFA of a test, the procedures for conducting a CFA, and the application of CFA to several important psychometric questions.

Summary

In this chapter, we have discussed the concept of test dimensionality and the way in which it is examined. We have discussed three core issues regarding test dimensionality: (1) the number of dimensions reflected in a set of test items, (2) the degree of association among a test's dimensions, and (3) the psychological meaning of a test dimension. These issues serve to differentiate three types of tests, which has important implications for the way a test is scored, evaluated, and used.

This chapter provided an overview of factor analysis—what it is and how it is used to examine test dimensionality. Although factor analysis is a highly advanced statistical procedure, we have provided a general discussion and illustration of the procedures. Interested readers can obtain more details from many available sources (e.g., Gorsuch, 1983; Meyers, Gamst, & Guarino, 2006).

The first four chapters of this book have provided the conceptual and statistical foundations for the remaining chapters. In the remaining chapters, we focus on core psychometric properties. Specifically, issues such as reliability and validity require familiarity with basic concepts and procedures such as variability, correlations, and dimensionality. We will turn next to reliability.

Suggested Readings

For a more extensive introduction to factor analysis:

Meyers, L. S., Gamst, G., & Guarino, A. (2006). *Applied multivariate research: Design and interpretation* (chaps. 12A–13B). Thousand Oaks, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

For a detailed technical discussion of the procedures:

Gorsuch, R. L. (1983). *Factor analysis.* Hillsdale, NJ: Lawrence Erlbaum.

For commentary and recommendations on common tendencies in the use of factor analysis:

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

For a detailed but concise discussion of the use of factor analysis in determining test score dimensionality:

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures.* Thousand Oaks, CA: Sage.

For recommendations on the use of factor analysis in scale development:

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299.