



Convenience samples of college students and research reproducibility



Robert A. Peterson ^{a,*}, Dwight R. Merunka ^{b,c,1}

^a University of Texas, United States

^b Paul Cézanne University Aix-Marseille, IAE Aix-en-Provence, Cergam, Rue de la Quille, Puyricard 13540, France

^c Euromed Management, Domaine de Luminy, BP 921, Marseille cedex 9 13288, France

ARTICLE INFO

Article history:

Received 24 November 2010

Received in revised form 7 September 2011

Accepted 8 December 2011

Available online 31 August 2013

Keywords:

Replication research

Student samples

Convenience samples

Research reproducibility

ABSTRACT

Tests of theory in marketing and consumer behavior research are frequently based on convenience samples of undergraduate college students. In a study of business-related ethicality, analysis of data from four dozen convenience samples of undergraduate business students revealed significant differences in means, variances, intercorrelations, and path parameters across the samples. Depending on the particular convenience sample used, relationships between variables and constructs were positive or negative and statistically significant or insignificant. The present research empirically documents, for the first time, the uncertainty created by using convenience samples of college students as research subjects. Only through empirical replications can researchers pragmatically assess the reliability, validity, and generalizability of research findings.

© 2013 Elsevier Inc. All rights reserved.

I challenge using students—and particularly students from a single department or school—as a sample population from which to generalize about consumer behavior, or even just for testing theory.

[— John Liefeld (2003, p. 12)]

This is not to say that findings based on students are always wrong. It is only to say that findings based on students are always suspect. Our findings would be substantially more credible if students were not so often the first and only choice.

[— William Wells (1993, p. 492)]

1. Introduction

One of the most contentious issues in consumer behavior research, and social science research generally, is the use of convenience samples of undergraduate college students as subjects in behavioral investigations. College students increasingly seem to be the subjects of choice in social psychology and consumer behavior research. To illustrate,

Peterson (2001) reports that college students constituted 86% of the research subjects in empirical studies appearing in Volume 26 of *Journal of Consumer Research*, whereas Simonson, Carmon, Dhar, Drolet, and Nowlis (2001) report that 75% of the research subjects in *Journal of Consumer Research* and *Journal of Marketing Research* articles were college students.

Arguments for and against the use of college students as research subjects have tended to focus on whether results obtained from such subjects are generalizable to non-student populations. Researchers such as Kardes (1996) and Lucas (2003) have argued that college students are appropriate research subjects when the research emphasis is on basic psychological processes or the theory tested links to human behaviors independent of sample characteristics. According to Berkowitz and Donnerstein (1982, p. 249), the “meaning the subjects assign to the situation they are in and the behavior they are carrying out plays a greater part in determining the generalizability of an experiment's outcome than does the sample's demographic representativeness.” However, other researchers, such as Sears (1986) and Wintre, North, and Sugar (2001), have expressed unease about the use of a narrow database of college students in behavioral research. In particular, Sears suggests that what is apparently “known” about humans is biased because college students tend to have stronger cognitive skills, less crystallized attitudes, more compliant behavior, and less stable peer group relationships than older adults.

During a nearly two-decade, highly cited dialogue, Calder and colleagues (Calder, Phillips, & Tybout, 1981, 1982, 1983; Calder & Tybout, 1999) debate Lynch (1982, 1983, 1999) regarding the need for external validity in consumer behavior research. This debate focuses on two types of empirical studies: effects application studies and theory application (or theoretical explanation) studies. Effects application

* Corresponding author at: Office of The Vice President for Research, The University of Texas at Austin, 1 University Station, Austin, Texas 78713, United States. Tel.: +1 512 471 9438; fax: +1 512 471 1035.

E-mail addresses: rap@mail.utexas.edu (R.A. Peterson), dwight.merunka@iae-aix.com (D.R. Merunka).

¹ Tel.: +33 662 154 675; fax: +33 442 280 800.

studies focus on substantive generalizability and questions of whether college students represent some larger population sufficiently well (e.g., consumers, managers) to warrant inferences from their responses to the larger population. In general, the scientific community seems to agree that effects application studies need samples that are more representative than college students of some relevant, “real-world” population (Bello, Leung, Radebaugh, Tung, & van Witteloostuijn, 2009; Lucas, 2003; Peterson, 2001).

Disagreement arises, however, regarding whether theory application studies require representative samples as a matter of methodological generalizability. The argument for using nonrepresentative samples, as Mook (1983, p. 384) articulates succinctly, is that “Representativeness of sample is of vital importance for certain purposes, such as survey research. For other purposes, it is a trivial issue.” “Other purposes” include research that aims to draw conclusions about theory rather than about a population. If the focus of research is theoretical, Mook (1983) argues that the makeup of a sample does not matter. Consequently, college students, or any other research participants, qualify as research subjects for fundamental research and theory testing (Bello et al., 2009; Mook, 1983; Pernice, van der Veer, Ommundsen, & Larsen, 2008).

However, even if theory testing is the study purpose, few researchers using convenience samples of college students appear to recognize that their investigation possesses the characteristics of a limited laboratory test that cannot generalize to other samples. Consider recently published research using convenience samples of students in *Journal of Business Research* (2009), *Journal of Consumer Research* (2009) and *International Journal of Research in Marketing* (2008 and 2009) and which formulated and tested hypotheses. Of 60 articles that contained 131 different theory-based studies conducted using convenience samples of college students, 38 (63%) ignored the sample usage in the discussion or conclusion section (e.g., Park & Lee, 2009; Yagci, Biswas, & Dutta, 2009; Yuksel & Mryteza, 2009). Conclusions typically focused on “people,” “participants,” “individuals,” “customers,” or “consumers.”

College students may be appropriate research subjects in certain situations, especially if they represent a population of interest. For example, business students are future business leaders, which may make them appropriate for studies in this domain (e.g., Abdolmohammadi, Gabhart, & Reeves, 1997; Ahmed, Chung, & Eichenseher, 2003; Borkowski & Ugras, 1998). College students may also be a key target market (e.g., Megehee, 2009) or used to replicate a prior study employing a student sample.

College student subjects might enhance research validity because of their apparent homogeneity. They tend to be homogeneous on dimensions such as age and education (which tend to influence attitudes), as well as possess weak self-definitions, high egocentrism, and a strong need for peer approval (Sears, 1986). Such homogeneity intuitively decreases variability in measurements and, ceteris paribus, increases the likelihood of rejecting a null hypothesis of no difference (Lynch, 1982, 1983), which in turn increases the probability of identifying theory violations when a theory is false (Lucas, 2003). Furthermore, such apparent homogeneity makes college student samples easier to compare than other groups of people because of their demographic and psychographic characteristics. This comparability premise may justify the choice of college student samples to represent different cultures in cross-cultural research (e.g., Aaker & Sengupta, 2000; Mikhailitchenko, Javalgi, Mikhailitchenko, & Laroche, 2009).

However, to deem college student samples acceptable or even recommended for theory-based behavioral research, the issue is not generalizability to other populations (e.g., general consumers) but generalizability to other college student samples. Assuming agreement that findings from a convenience sample of college students in one university (who often are drawn from only one class) do not generalize to different populations (e.g., managers, general consumers), the question that remains is whether researchers can

replicate the findings from a convenience sample of college students under similar research conditions.

Therefore, a key issue relating to the use of convenience samples of students to test theory is reproducibility, or whether under similar conditions, the findings replicate. In brief, the present study considers the extent to which research findings obtained from a convenience sample of college students at a single college or university can be replicated with convenience samples of college students in other colleges or universities. Although researchers employing college student samples may conduct pretests and validation and cross-validation studies, the studies typically employ a single subject pool from one department, college, or university.

Despite widespread concerns surrounding the use of convenience samples of college students for theory testing (Ferber, 1977; Peterson, 2001), the authors could not find any study that offered convincing empirical evidence regarding the negative consequences for research conclusions drawn from them. Nor have proponents of using convenience samples of college students offered convincing empirical evidence regarding their benefits (other than cost and convenience). Rather, proponents have simply argued that because they study theoretical effects, not personal characteristics, the convenience sampling issue is moot. An additional argument—without empirical support—is that because college student samples are homogeneous on many dimensions, theory testing with these samples might be more valid than testing with nonstudents because of the reduction in measurement variability.

For example, in a study of information incongruity, Aaker and Sengupta (2000) justify the choice of student subjects from undergraduate programs in major universities because of an alleged high degree of similarity on demographic and psychographic dimensions. Strizhakova, Coulter, and Price (2008) analyze data on samples of college students because the students “exhibited” a high degree of homogeneity and could be compared with minimal extraneous biases across multiple cultural sites. Finally, numerous articles reporting the results of consumer behavior research present multiple experiments that are cumulative in their results and findings. Thus, building on results from one convenience sample of students to another implies inter-sample homogeneity (e.g., Lalwani, 2009 used five different convenience samples of undergraduate students in five consecutive experiments). The present research investigates empirically the extent to which a particular convenience sample of college students produces research findings identical to, or at least consistent with, research findings from similar convenience samples of college students. Peterson's (2001) results clearly augur against generalizing from college students to nonstudents, but the possibility of generalizing from a “typical” convenience sample of college students to a larger body of convenience samples of college students remains unclear. If the results from a “typical” convenience sample of college students do not generalize to a larger body of college students, then the rationale for using convenience samples of college students as research subjects for theory testing, because of their homogeneity or generalizability, is suspect.

2. Theory testing

To investigate the use of convenience samples of college students to test theories through formal hypotheses, this study focuses on the attitudinal domain of business ethics. Students are often participants in investigations in this domain, and both precedents and a rationale exist for studying undergraduate business students. For example, undergraduate business students have often been studied because they represent prospective managers (Preble & Reichel, 1988; Stevenson & Bodkin, 1998) or business executives (Ahmed et al., 2003; Jones & Gautschi, 1988). Further, by employing a relatively homogeneous group of individuals, minimizing possible contaminants (e.g., family status, work experience, academic major) of perceptions of business ethics

or even ethical behavior becomes possible. Therefore, the domain for the present research favors, *ceteris paribus*, response homogeneity across samples of business students. In the present context, theory testing focuses on two constructs, attitude toward business ethics and attitude toward capitalism, and two variables hypothesized to relate to these attitudes, gender and religiosity.

2.1. Gender and ethicality

Many business ethics studies have found that women show higher ethical standards and behaviors than men. Borkowski and Ugras (1998) conducted a quantitative review of 47 empirical studies of the relationship between gender and ethics. After conducting meta-analyses of both statistical significance levels and effect sizes, the authors concluded that “the null hypothesis of no relationship between gender and ethical behavior can be rejected” (Borkowski & Ugras, 1998, p. 1124); women showed more ethicality than men. Similarly, Franke, Crown, and Spake (1997) conducted a meta-analysis to investigate the role that gender plays in perceptions of ethical decision making and found gender differences in pre-career (student) samples, with women having higher ethical standards than men. Following a qualitative literature review, Kennedy and Lawton (1996, p. 904) conclude that although some studies “have shown little or no difference between males and females ... none have found higher standards for males than females.” Therefore,

H1. Female business students show a higher level of business-related ethicality than do male business students.

2.2. Religiosity and ethicality

Studies of the relationship between religiosity and ethics have dealt with issues such as cheating by students (Allmon, Page, & Roberts, 2000; Barnett, Bass, & Brown, 1996), environmentalism (Wolkomer, Futreal, Woodrum, & Hobau, 1997) and insider trading (Terpstra, Rozell, & Robinson, 1993). These studies found positive relationships between degree of religiosity and ethical attitudes. Likewise, the relationship between degree of religiosity and business-related ethicality generally appears positive (e.g., Conroy & Emerson, 2004; Ibrahim, Howard, & Angelidis, 2008; Kurpis, Beqiri, & Helgeson, 2008; Wong, 2008). Thus, the extant literature suggests the following hypothesis:

H2. Business students show a positive relationship between degree of religiosity and business-related ethicality.

2.3. Capitalism and ethicality

Definitions of capitalism include two main dimensions: private ownership of property and consensual exchanges in a free market (Bishop, 2000). Founded on the imperatives of individual rights and political and religious freedom, capitalism considers private property a moral right and beneficial because property supports other moral social institutions such as freedom, democracy, and law enforcement (Usher, 2000). In turn, the free exchange of goods and services is a form of freedom and beneficial because such exchanges create price and thus provide information (Friedman, 1962). Most assessments consider capitalism and competitive market situations morally superior to both feudalism and Marxism, and the values of capitalism are associated with Protestant social ethics (Buchholz, 1983; Carr, 2003). Therefore, people who favor capitalism and its associated values (e.g., freedom, democracy, equity, private ownership, defense and protection of property rights) should possess positive attitudes toward business-related ethics. Consequently:

H3. Business students show a positive relationship between attitude toward capitalism and business-related ethicality.

3. Empirical investigation

For this investigation into whether research results from convenience samples of college students are reproducible, a judgmentally representative sample of 64 four-year business schools in the United States was selected and a faculty member contacted in each school. In each of the schools the faculty member or an assistant administered a questionnaire to undergraduate business students in an in-class setting. (The in-class setting controlled for possible noise in data collection because the data collection environments were relatively similar.) The questionnaires did not take more than a few minutes to administer, and the faculty member or assistant received an honorarium of \$20 as a token of appreciation. A pilot test of the questionnaire with a separate sample of business students provided a qualitative evaluation of its comprehensibility and an assessment of administrative ease.

All samples with 30 usable questionnaires or more were included in the analysis. In total, 49 samples represented business schools located in 30 states ranging from Maine to California, Washington to Georgia, and Minnesota to Texas. Comparative data from the United States Census Bureau and the Association to Advance Collegiate Schools of Business suggested that the gender and age characteristics of the sample as a whole were consistent with those of undergraduate (four-year) college students nationally, as well as undergraduate business students generally. The ratio of male to female students (approximately 50–50) was relatively consistent across the 49 samples. The total sample consisted of 2761 undergraduate business students, with an average of 56 students per sample.

The questionnaire consisted of 27 rating scales that measured attitudes, as well as three items (academic classification, citizenship, and major field of study) to ensure that the sample was limited to U.S. undergraduate business students, and five demographic questions (gender, age, whether employed, years at university, and religiosity). Each of the 27 rating scales consisted of a declarative statement and six response categories ranging from “strongly agree” (“1”) to “strongly disagree” (“6”). These 27 rating scales provided the input for two multi-item attitude scales, attitude toward business ethics, *Aethics*, and attitude toward capitalism, *Acapitalism*. Each attitude scale consisted of four items. Gender was a dichotomous item (male = 0, female = 1). The religiosity measure included a three-category item, “very religious,” “somewhat religious,” and “not very religious.” The “very” and “somewhat” religious categories were collapsed to create a dichotomous variable (very + somewhat = 1, not very = 0).

3.1. Measurement properties

Exploratory factor analyses conducted on the items comprising the two scales for the total sample led to the elimination of one item per scale.² A confirmatory factor analysis conducted with the total sample demonstrated the reliability (Jöreskog’s rho greater than .8) and discriminant validity (average variances extracted [AVE] greater than .5) of the two scales. As additional support for discriminant validity, the square root of the AVEs exceeded the correlation between the constructs, and cross-loadings were small compared with loadings. Measurement invariance for the *Aethics* scale was assessed through multi-group analyses for pairs of groups based on gender (female and male) and religiosity level (very and somewhat versus not very). Measurement invariance existed across gender and religiosity.

Confirmatory factor analyses conducted on each of the 49 samples helped determine if the two scales were appropriate for every sample and indicated reliability and discriminant validity (criteria: Jöreskog’s rho > .8, AVE > .5). The AVE for the *Aethics* scale ranged from .24 to .79, whereas that for the *Acapitalism* scale ranged from .20 to .76. Samples with AVE less than .5 were excluded from further analysis;

² Scale items are available on request.

this procedure included five samples based on results for the *Aethics* scale and three samples based on results for the *Acapitalism* scale (one excluded sample was common to both scales). Consequently, the evaluations of H1 and H2 used 44 samples, whereas the H3 evaluation used 42 samples.

3.2. Descriptive analysis

Each of the convenience samples of college students is an independent replication, so determining whether statistics derived from any one sample differed from statistics derived from any other sample is instructive. Figs. 1 and 2 graph the means and variances of the 44 samples for the *Aethics* scale and the 42 samples for the *Acapitalism* scale, respectively. As these figures show, considerable variability exists across both the means and variances of the samples. Sample means range from 5.0 to 8.4 for the *Aethics* scale and from 5.3 to 10.7 for the *Acapitalism* scale, whereas sample variances range, respectively, from 5.3 to 17.0 and from 4.4 to 12.1 for the two scales. Although the sample variances show wider ranges than the sample means, statistically significant differences mark the extreme means and extreme variances at the .05 level. Each of the largest ten (six) sample variances is significantly different from each of the smallest ten (seven) sample variances for *Aethics* (*Acapitalism*). Each of the largest nine (thirteen) sample means is significantly different from each of the smallest nine (thirteen) sample means for *Aethics* (*Acapitalism*). Thus the responses are not uniformly homogeneous across the samples, and many of the samples differ considerably in terms of their attitudes toward business-related ethics and capitalism.

To further investigate heterogeneity in responses, the means of the *Aethics* and *Acapitalism* scales of each sample are respectively compared to those of every other sample on a pairwise basis. This process is equivalent to the situation where a researcher seeks to replicate a finding from one sample on another “equivalent” sample. A total of 924 pairwise comparisons were conducted for the means of the *Aethics* scale and 861 pairwise comparisons for the means of the *Acapitalism* scale. Differences between the pairs of the *Aethics* scale means are statistically significant in 17.1% of the comparisons (at $p < .05$). Differences between the pairs of the *Acapitalism* scale means are statistically significant in 30.6% of the comparisons (at $p < .05$). For some of the samples significant differences in the *Aethics* scale means with more than 50% of the other samples emerged.

Even if the samples used had been selected because of their homogeneity (all were undergraduate business student samples from U.S. universities and consisted of American students), some variations across

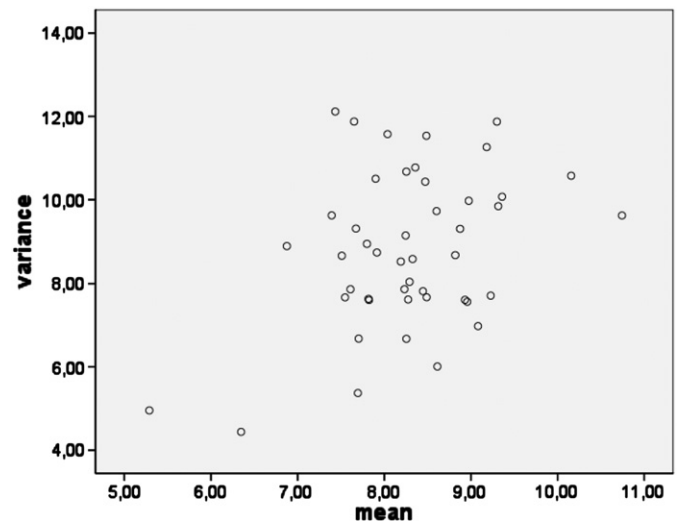


Fig. 2. Means and variances of the *Acapitalism* scale in 42 convenience samples of business students.

samples may explain the differences obtained. For example, religiosity varies across regions in the United States. Therefore, a median split was performed on the samples based on average religiosity within each sample. Across the low-religiosity samples (mean < 2.2 on a 1 to 3 scale), significant differences ($p < .05$) across pairs of samples in 20.2% of the comparisons for the *Aethics* scale exist. Across high-religiosity samples (mean > 2.2), there are significant differences ($p < .05$) in 15.3% of the comparisons for the *Aethics* scale. The corresponding percentages are respectively 27.5% and 32.1% for the *Acapitalism* scale. Therefore, within homogeneous religiosity samples, observed differences across samples remain the same as observed for the total sample.

4. Hypothesis tests

The three hypotheses relate (1) gender, (2) degree of religiosity, and (3) attitude toward capitalism to attitude toward business ethics. Evaluations of the hypotheses are first executed within each sample, with the results then compared across samples. Because the research focus is not on substantive issues but rather on relative differences in the results across samples, each hypothesis was first independently evaluated by means of bivariate correlation coefficients to provide baseline information. (Note that in each analysis the correlation coefficients are independent of sample sizes; point-biserial correlation coefficients are used for the gender and religiosity analyses.) This analysis was followed by a multivariate test of the relative differences in results across the samples using partial least squares (PLS) path modeling.

4.1. Gender and ethicality (H1)

Gender relates negatively to *Aethics* in seven samples (16%) at $p < .05$ (correlation coefficients range from $-.20$ to $-.47$). The negative coefficients indicate that women have higher *Aethics* scores in the seven samples than do men, in support of H1. However, 37 samples (84%) display no significant relationship between gender and *Aethics* (correlation coefficients vary from $-.22$ to $+.18$; 34 of the 44 correlation coefficients are negative, in directional support of H1). Thus, depending on the convenience sample employed, the hypothesized relationship between gender and *Aethics* is either insignificant or significant and either positive or negative. The correlation between gender and *Aethics* is negative ($r = -.08$) for the total sample, significant at $p < .01$, indicating higher *Aethics* values for women than for men (H1 supported).

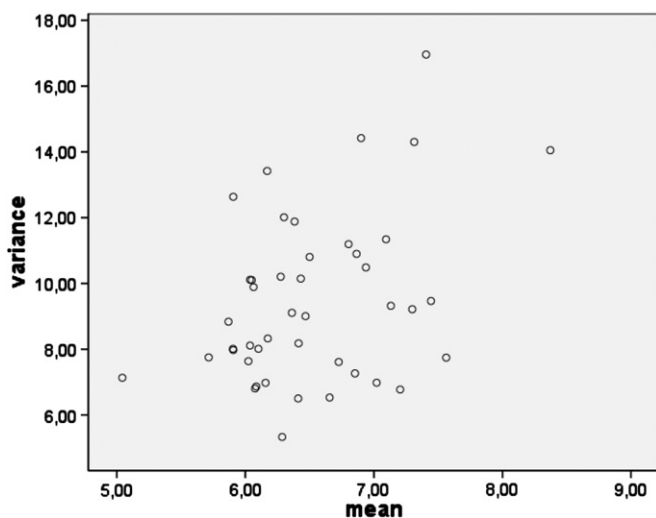


Fig. 1. Means and variances of the *Aethics* scale in 44 convenience samples of business students.

4.2. Religiosity and ethicality (H2)

Religiosity does not relate to *Aethics* in 42 of the 44 samples. Across the 44 samples, correlation coefficients range from $-.37$ to $+.24$, with 22 coefficients being positive and 22 being negative. The two significant coefficients ($p < .05$) are negative ($-.25$ and $-.37$); greater reported religiosity is related to greater business-related ethicality. Thus, H2 receives support in two of the 44 samples. At the aggregate level, reported religiosity and *Aethics* are independent ($r = .01$, $p > .05$), which does not provide support for H2.

4.3. Attitude toward capitalism and attitude toward business ethics (H3)

Attitude toward capitalism relates positively and significantly ($p < .05$) to attitude toward business ethics in 12 of the 42 samples (29%), with correlations ranging from $+.23$ to $+.48$ in support of H3. In the remaining 30 samples (71%), the correlation coefficients vary from $-.10$ to $+.27$, but their lack of significance does not support H3. However, 35 of the 42 correlation coefficients are directionally supportive (i.e., positive). The correlation between *Acapitalism* and *Aethics* is $+.18$ ($p < .01$) for the total sample; H3 therefore receives support at the aggregate level.

4.4. Path analysis

A multivariate test of the generalizability of results across the samples consists of evaluating a model that combines the three hypotheses and therefore the joint effects of gender, religiosity, and attitude toward capitalism on attitude toward business ethics. The model (Fig. 3) relates gender, religiosity, and *Acapitalism* to *Aethics*.

The sample sizes made PLS path modeling appropriate (Tenenhaus, Vinzi, Chatelin, & Lauro, 2005). Compared to covariance-based Structural Equation Modeling (SEM), PLS path analysis is more robust and deals well with relatively small sample sizes (Sosik, Kahai, & Piovoso, 2009). PLS is well suited for predicting structural relationships and provides robust estimates of a structural model (Hair, Ringle, & Sarstedt, 2011). The primary goal when applying PLS path analysis is to examine how well gender, religiosity, and *Acapitalism* explain the endogenous latent construct (*Aethics*) by examining R^2 values and the respective structural path coefficients.

Parameter estimates and significance testing rely on a bootstrapping procedure ($n = 200$). The results indicate significant relationships for 52% (22) of the samples based on model R^2 values ($p < .05$). R^2 values vary from .09 to .38, with R^2 values above .15 being statistically significant.

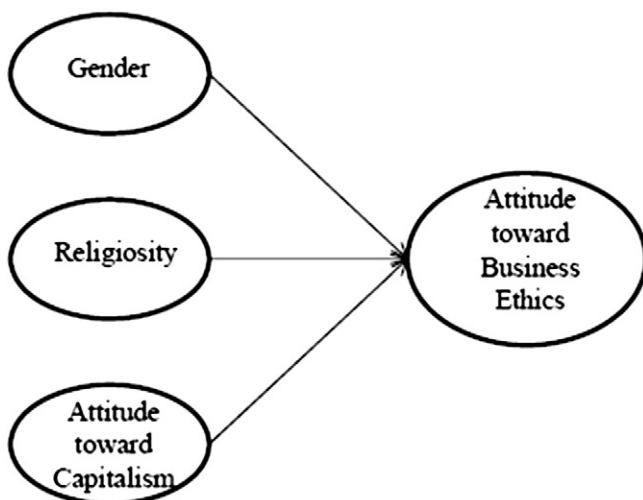


Fig. 3. Model of attitude towards business ethics.

The path coefficient for gender is significant for 24% of the samples (10/42) and varies from $-.22$ to $-.46$; these samples support H1. Religiosity relates significantly to *Aethics* in four samples (10%), with two positive and two negative path coefficients. *Acapitalism* relates positively to *Aethics* in 17 samples (40%), with path coefficients ranging from $+.25$ to $+.55$ ($p < .05$) in support of H3. However, H3 does not receive support in the remaining 25 samples. At the aggregate level, R^2 is .05 ($p < .01$), and both gender and *Acapitalism* relate significantly to *Aethics*. The path coefficient is negative for gender ($\beta = -.13$, $p < .01$), in support of H1. The path coefficient for *Acapitalism* is positive ($\beta = +.22$, $p < .01$), in support of H3. However, religiosity is not related to *Aethics* ($\beta = -.009$), so H2 does not receive support. Table 1 summarizes the results of hypotheses testing at the sample level.

These results empirically demonstrate the potential for inconsistency in research outcomes when using convenience samples of college students in theory testing, even when the choice of students appears theoretically justified (i.e., business students are future business executives whose ethical attitudes are of interest). Depending on the convenience sample examined, the relationships observed were positive or negative, statistically significant or insignificant. Consequently, in this instance the results obtained from convenience samples of college students demand caution, even for theory-testing research in which these samples seem appropriate.

5. The need for research replications

This investigation did not employ an experimental research design and consequently could not address all the issues posed in the Calder–Lynch debate. Even so, the results are disconcerting. Substantively significant differences, in both direction and magnitude, emerge across some four dozen convenience samples of college students with respect to scale means and variances, intercorrelations, and structural relationships. Because the convenience samples in the study are quantitatively and qualitatively similar, as well as similar to the types of samples frequently used in behavioral research in the United States, the results suggest that the use of any particular convenience sample may or may not produce results that comport with results obtained from any other particular convenience sample, or even the aggregate sample as a whole. The present study even had to exclude seven original samples when evaluating the hypotheses because the scales for these samples did not meet standard reliability and discriminant validity criteria. Regardless of whether observed sample differences are due to measurement unreliability, small sample sizes, convenience sampling, or some unknown demand artifacts, they reflect an empirical phenomenon analogous to randomly sampling a distribution of convenience samples.

One possible limitation of the study is the average size of the samples of college students. Given that the unit of analysis was a sample of college students rather than an individual student, the goal was to collect data from many samples rather than to obtain large sample sizes.

Table 1
Results of the tests of hypotheses on all 42 samples.

Univariate analysis	Hypothesis testing	Range of coefficients ^a	Positive coefficients	Negative coefficients	Significant relationships
Gender	H1	$-.47, +.18$	23%	77%	16%
Religiosity	H2	$-.37, +.24$	50%	50%	3%
Acapitalism	H3	$-.10, +.48$	83%	17%	29%
Multivariate analysis		Range of coefficients ^b			
Gender	H1	$-.46, +.40$	7%	93%	24%
Religiosity	H2	$-.33, +.32$	41%	59%	10%
Acapitalism	H3	$-.15, +.55$	98%	2%	40%

^a Correlation coefficients.

^b Path coefficients.

Doing so resulted in an average sample size of 56 (which is not unusual in consumer research). In spite of the bootstrap procedure used, the relatively small sample sizes may have accounted for some of the differences observed between samples. Thus replications of the present research with larger sample sizes are advisable. Another potential limitation is related to possible model mis-specification. Gender, religiosity, and attitude towards capitalism predicted attitude toward business ethics. These variables have previously been studied but are far from explaining all variation in ethicality since the maximum R^2 obtained is .38. Omitted variables could have influenced the results in spite of the demographic homogeneity of the samples.

Regardless of these possible limitations, study results empirically corroborate Epstein's (1979, 1980) conclusion that "the very nature of the paradigm of the single-session experiment [study] is such that very few findings, no matter what their level of statistical significance, are apt to be replicable" (Epstein, 1980, p. 790). Yet as Epstein (1980, p. 796) opines, "There is no more fundamental requirement in science than that the replicability of findings be established." Consequently, the primary takeaway from the present research is that marketing and consumer behavior researchers must conduct extensive replications of their own studies, as well as the studies of others, especially those that use convenience samples of college students, even if the purpose of the research is theory testing and in spite of the belief that student samples are very homogeneous.

If the results from college student convenience samples are not reproducible (Evanschitzky, Baumgarth, Hubbard, & Armstrong, 2007; Hubbard & Armstrong, 1994), other, independent samples of students and nonstudents are necessary. Even the present study demands replication, with undergraduate business students drawn from other colleges and universities, different variables and constructs, and different methodologies. The argument that the characteristics of research subjects are irrelevant for theoretical research or that college students are homogeneous research subjects rings hollow in light of the empirical inconsistencies observed among the wide range of convenience samples herein. The present study vividly illustrates the possibility of type-I errors when research subjects consist of only a single convenience sample.

Calls for replications are commonplace (e.g., Easley, Madden, & Dunn, 2000; Hubbard & Vetter, 1996; Hunter, 2001; Monroe, 1992a,b; Wells, 2001), but unfortunately seem to go unheeded. The obvious efficiency and cost effectiveness of using college students as subjects seem to dominate issues of validity, but the cost of invalid inferences must be considered also. This empirical confirmation that the use of convenience samples of college students as research subjects is fraught with reliability, validity, and generalizability problems demands a full embrace of replications. As Lindsay and Ehrenberg (1993, p. 236) conclude, "If a study is worth doing at all, it's worth doing twice." Only replications can assess the reliability, validity, and generalizability of research findings pragmatically. Only replications can reduce the uncertainty associated with any particular set of research results or sample.

Unfortunately, how many and what types of replications are necessary remain unclear. Replications can vary from faithful duplications to close replications to differentiated and conceptual studies (Easley et al., 2000; Lindsay & Ehrenberg, 1993). Holdout samples and cross-validation samples do not qualify as replications, nor do bootstrapping and jackknifing. These methods only reflect the sampling process used; the underlying sample is constant.

Publication of an empirical investigation in a top-tier journal should require a minimum of two within-study replications, and at least one of the replications should use research subjects qualitatively different from those employed in the validation study (i.e., a different subject pool). This requirement is not too onerous. Haslam and McGarty (2001) find that the average number of experiments per study in articles in *Journal of Personality and Social Psychology* increased from slightly more than one in 1968 to approximately three in 1998. In the 2009 volume of *Journal of Consumer Research*, articles reporting empirical study results (excluding qualitative, meta-analyses, and purchase-

history electronic database studies) relied on an average of 3.5 samples—although most of these samples consisted of college students drawn from single-source subject pools. Adding replications to a research agenda would not seem to be an unreasonable burden.

To determine the number and type of external or independent replications of studies, the Bayesian framework offered by Raman (1994) or the approach of Hunter (2001) might serve as guides. The methodology proposed by Farley, Lehmann, and Mann (1998) for designing studies also might prove useful in determining the types of replications depending on the goal (i.e., reducing random error, validating or confirming conclusions, extending conclusions, and/or detecting bias), which seem especially relevant if the object of the replication used a convenience sample of college students. Despite such methodological progress, the number and type of replications require more work and further investigation.

6. Final thoughts

Marketing and consumer behavior researchers frequently violate the scientific canon that statistical inferences should be limited to the populations from which samples are drawn. The present research reveals that statistical inferences drawn from convenience samples of business students do not even generalize to a business student population. This finding prompts the following recommendation: Every manuscript submitted to a top-tier journal for publication consideration that reports empirically based research should contain an explicit statement justifying the theoretical relevance of the subjects employed to test the specific research questions for the population of interest. This statement should be an integral part of the conceptual foundation or research methodology section of any manuscript, distinct from the limitations section. That is, authors should theoretically justify the sample used prior to conducting their research, not apologize for or ignore the usage subsequently. For example, Balabanis and Diamantopoulos (2004, p. 84) explicitly stated in their data collection section, "in this particular study, the use of a student sample would have been inappropriate ... if the external validity of our study was to be safeguarded."

In the short run, requiring an explicit, theoretical justification for a sample emphasizes and reinforces the importance of using appropriate research subjects. In the long run, such a justification should lead to more research replications, more research that is reproducible, more relevant research, and more generalizable knowledge.

Finally, the results of the present study might imply some support for a shift in the traditional behavioral research paradigm, from induction to (strict) falsification. However, the disparate results observed across samples are as problematic for falsification-based research as they are for induction-based research, because the representativeness of any one specific sample remains unknown. To the extent that college students are not a homogeneous population, the argument that "any sample will do" to evaluate theoretical effects is specious, regardless of the research paradigm. At best, convenience samples of college student subjects can serve a useful function for identifying potential boundary conditions of a theory.

References

- Aaker, J. L., & Sengupta, J. (2000). Additivity versus attenuation: The role of culture in the resolution of information incongruity. *Journal of Consumer Psychology*, 9, 67–82.
- Abdolmohammadi, M. J., Gabhart, D. R. L., & Reeves, M. F. (1997). Ethical cognition of business students individually and in groups. *Journal of Business Ethics*, 16, 1717–1725.
- Ahmed, M. M., Chung, K. Y., & Eichenseher, J. W. (2003). Business students' perception of ethics and moral judgment: A cross-cultural study. *Journal of Business Ethics*, 43, 89–102.
- Allmon, D. E., Page, D., & Roberts, R. (2000). Determinants of perceptions of cheating: Ethical orientation, personality, and demographics. *Journal of Business Ethics*, 23, 411–422.

- Balabanis, G., & Diamantopoulos, A. (2004). Domestic country bias, country-of-origin effects, and consumer ethnocentrism: A multidimensional unfolding approach. *Journal of the Academy of Marketing Science*, 32, 80–95.
- Barnett, T., Bass, K., & Brown, G. (1996). Ethical ideology, and intentions to report a peer's wrongdoing. *Journal of Business Ethics*, 15, 1161–1174.
- Bayarri, M. J., & Mayoral, A.M. (2002). Bayesian design of "successful" replications. *American Statistician*, 56, 207–214.
- Bello, D., Leung, K., Radebaugh, L., Tung, R. L., & van Witteloostuijn, A. (2009). From the editors: Student samples in international business research. *Journal of International Business Studies*, 40, 361–364.
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep. *American Psychologist*, 37, 245–257.
- Bishop, J.D. (2000). Ethics and capitalism: A guide to the issues. In J.D. Bishop (Ed.), *Ethics and capitalism* (pp. 3–48). Toronto, Canada: University of Toronto Press.
- Borkowski, S.C., & Ugras, Y. J. (1998). Business students and ethics: A meta-analysis. *Journal of Business Ethics*, 17, 1117–1127.
- Buchholz, R. A. (1983). The protestant ethic as an ideological justification of capitalism. *Journal of Business Ethics*, 2, 51–60.
- Calder, B. J., Phillips, L. W., & Tybout, A.M. (1981). Designing research for application. *Journal of Consumer Research*, 8, 197–207.
- Calder, B. J., Phillips, L. W., & Tybout, A.M. (1982). The concept of external validity. *Journal of Consumer Research*, 9, 240–244.
- Calder, B. J., Phillips, L. W., & Tybout, A.M. (1983). Beyond external validity. *Journal of Consumer Research*, 10, 112–114.
- Calder, B. J., & Tybout, A.M. (1999). A vision of theory, research, and the future of business schools. *Journal of the Academy of Marketing Science*, 27, 359–366.
- Carr, P. (2003). Revisiting the protestant ethic and the spirit of capitalism: Understanding the relationship between ethics and enterprise. *Journal of Business Ethics*, 47, 7–16.
- Conroy, S. J., & Emerson, T. L. N. (2004). Business ethics and religion: Religiosity as a predictor of ethical awareness among students. *Journal of Business Ethics*, 50, 383–396.
- Easley, R. W., Madden, C. S., & Dunn, M. G. (2000). Conducting marketing science: The role of replication in the research process. *Journal of Business Research*, 48, 83–92.
- Epstein, S. (1979). The stability of behavior I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097–1126.
- Epstein, S. (1980). The stability of behavior II. Implications for psychological research. *American Psychologist*, 35, 790–806.
- Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60, 411–415.
- Farley, J. U., Lehmann, D. R., & Mann, L. H. (1998). Designing the next study for maximum impact. *Journal of Marketing Research*, 35, 496–501.
- Ferber, R. (1977). Research by convenience. *Journal of Consumer Research*, 4, 57–58.
- Franke, G. R., Crown, D. F., & Spake, D. F. (1997). Gender differences in ethical perceptions of business practices: A social role theory perspective. *Journal of applied psychology*, 82, 920–934.
- Friedman, M. (1962). *Capitalism and freedom*. Chicago, IL: The University of Chicago Press.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19, 139–151.
- Haslam, S. A., & McGarty, C. (2001). A 100 years of certitude? School psychology, the experimental method, and the management of scientific uncertainty. *British Journal of Social Psychology*, 40, 1–21.
- Hubbard, R., & Armstrong, J. S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11, 233–248.
- Hubbard, R., & Vetter, D. E. (1996). An empirical comparison of published replication research in accounting, economics, finance, management, and marketing. *Journal of Business Research*, 35, 153–164.
- Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, 28, 149–158.
- Ibrahim, N., Howard, D., & Angelidis, T. (2008). The relationship between religiousness and corporate social responsibility orientation: Are there differences between business managers and students? *Journal of Business Ethics*, 78, 165–174.
- Jones, T. M., & Gauthier, F. H., III (1988). Will the ethics of business change? A survey of future executives. *Journal of Business Ethics*, 7, 231–248.
- Kardes, F. R. (1996). In defense of experimental consumer psychology. *Journal of Consumer Psychology*, 5, 279–296.
- Kennedy, E. J., & Lawton, L. (1996). The effects of social and moral integration on ethical standards: A comparison of American and Ukrainian business students. *Journal of Business Ethics*, 15, 901–911.
- Kurpis, L. V., Beqiri, M. S., & Helgeson, J. G. (2008). The effects of commitment to moral self-improvement and religiosity on ethics of business students. *Journal of Business Ethics*, 80, 447–463.
- Lalwani, A. K. (2009). The distinct influence of cognitive busyness and need for closure on cultural differences in socially desirable responding. *Journal of Consumer Research*, 36, 305–316.
- Liefeld, J. P. (2003). Consumer research in the land of Oz. *Market Research*, 15, 10–15.
- Lindsay, R. M., & Ehrenberg, A. S.C. (1993). The design of replicated studies. *American Statistician*, 47, 217–228.
- Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21, 236–253.
- Lynch, J. G., Jr. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9, 225–239.
- Lynch, J. G., Jr. (1983). The role of external validity in theoretical research. *Journal of Consumer Research*, 10, 109–111.
- Lynch, J. G., Jr. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27, 367–376.
- Megehee, C. M. (2009). Advertising time expansion, compression, and cognitive processing influences on consumer acceptance of message and brand. *Journal of Business Research*, 62(4), 420–431.
- Mikhailitchenko, A., Javalgi, R. G., Mikhailitchenko, G., & Laroche, M. (2009). Cross-cultural advertising communication: Visual imagery, brand familiarity, and brand recall. *Journal of Business Research*, 62(10), 931–938.
- Monroe, K. (1992a). On replications in consumer research: Part I. *Journal of Consumer Research*, 19, i–ii.
- Monroe, K. (1992b). On replications in consumer research: Part II. *Journal of Consumer Research*, 19, i–ii.
- Mook, D.G. (1983). In defense of external invalidity. *American Psychologist*, 38, 379–387.
- Park, C., & Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*, 62(1), 61–67.
- Pernice, R. E., van der Veer, K., Ommundsen, R., & Larsen, K. (2008). On use of student samples for scale construction. *Psychological Reports*, 102, 459–464.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28, 450–461.
- Preble, J. F., & Reichel, A. (1988). Attitudes toward business ethics of future managers in the U.S. and Israel. *Journal of Business Ethics*, 7, 941–949.
- Raman, K. (1994). Inductive inference and replications: A Bayesian perspective. *Journal of Consumer Research*, 20, 633–643.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.
- Simonson, I., Carmon, Z., Dhar, R., Drolet, A., & Nowlis, S. M. (2001). Consumer research: In search of identity. In S. T. Fiske, D. L. Schacter, & C. Zahn-Waxler (Eds.), *Annual Review of Psychology*, Vol. 52. (pp. 249–275) Palo Alto, CA: Annual Reviews.
- Sosik, J. J., Kahai, S. S., & Piovoso, M. J. (2009). Silver bullet or voodoo statistics? A primer for using partial least squares data analytic technique in group and organization research. *Group & Org. Management*, 34, (pp. 5–36).
- Stevenson, T. H., & Bodkin, C. D. (1998). A cross-national comparison of university students' perceptions regarding the ethics and acceptability of sales practices. *Journal of Business Ethics*, 17, 45–55.
- Strizhakova, Y., Coulter, R. A., & Price, L. L. (2008). The meanings of branded products: A cross national scale development and meaning assessment. *International Journal of Research in Marketing*, 25, 82–93.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y. -M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48, 159–205.
- Terpstra, D. E., Rozell, E. J., & Robinson, R. K. (1993). The influence of personality and demographic variables on ethical decisions related to insider trading. *Journal of Psychiatry*, 127, 375–389.
- Usher, D. (2000). The justification of private property. In J.D. Bishop (Ed.), *Ethics and capitalism* (pp. 49–80). Toronto, Canada: University of Toronto Press.
- Wells, W. D. (1993). Discovery-oriented consumer research. *Journal of Consumer Research*, 19, 489–504.
- Wells, W. D. (2001). The perils of N = 1. *Journal of Consumer Research*, 28, 494–498.
- Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology*, 42, 216–225.
- Wolkmer, M., Futreal, M., Woodrum, F., & Hobau, T. (1997). Substantive religious belief and environmentalism. *Social Science Quarterly*, 78, 96–108.
- Wong, H. (2008). Religiousness, love of money, and ethical attitudes of Malaysian evangelical Christians in business. *Journal of Business Ethics*, 81, 169–191.
- Yagci, M. I., Biswas, A., & Dutta, S. (2009). Effects of comparative advertising format on consumer responses: The moderating effects of brand image and attribute relevance. *Journal of Business Research*, 62(8), 768–774.
- Yuksel, U., & Mryteza, V. (2009). An evaluation of strategic responses to consumer boycotts. *Journal of Business Research*, 62(2), 248–259.