SOCRATES, STRATEGY, AND STRUCTURAL MODELING:

MEASUREMENT ERROR AND STRATEGIC MANAGEMENT RESEARCH

Brian K. Boyd

Corporate MBA Program
W. P. Carey School of Business
Arizona State University
Tempe, AZ 85287-4111
(480) 965-4781 (voice)
(480) 727-7494 (fax)
briankboyd@asu.edu
http://www.briankboyd.com

Steve Gove

Management / Marketing Department
University of Dayton
Miriam Hall 713
300 College Park
Dayton, OH 45469-2271
(937) 229-2239 (voice)
(937) 229-3788 (fax)
gove@udayton.edu

Michael A. Hitt

Management Department
Mays Business School
Texas A&M University
College Station, TX 77843-4221
(979) 458-3393
(979) 845-3420
mhitt@cgsb.tamu.edu

SOCRATES, STRATEGY, AND STRUCTURAL MODELING:

MEASUREMENT ERROR AND STRATEGIC MANAGEMENT RESEARCH

ABSTRACT

Strategic management research has been characterized as placing less emphasis on construct measurement than other management subfields. In this study, we document the state of the art of measurement in strategic management research, and discuss the implications for interpreting the results of research in this field. To illustrate the consequences of measurement error, we revisit the debate on the causes of diversification (Amihud & Lev, 1982). Our research suggests that the divergent findings between studies on this topic are largely the result of measurement error, and that prior work has underestimated the true effect size between variables. To assess the breadth of the issue in the discipline, we conducted a content analysis of empirical strategic management articles published in leading journals in the period of 1998-2000 and found that many studies do not discuss reliability and validity issues, and typically rely on single-indicator measures. Additionally, studies rarely address the problems of attenuation due to measurement error. We close with a discussion of the implications for future research and for interpreting prior work in strategic management.

Key Words: Measurement, research design, agency theory, diversification, corporate governance

Book 7 of *The Republic* presents what is probably the most widely known parable of Socrates – the shadows on the cave wall. As he said to Glaucon:

> Behold! human beings living in an underground den, which has a mouth open toward the light and reaching all along the den; here they have been since childhood, and have their legs and necks chained so that they cannot move, and can only see before them, being prevented by the chains from turning round their heads. Above and behind them is a fire blazing at a distance, and between the fire and the prisoners there is a raised way; and you will see, if you look, a low wall built along the way, like the screen which marionette-players have in front of them, over which they show the puppets
>
> I see.
>
> And do you see, I said, men passing along the wall carrying all sorts of vessels, and statues and figures of animals made of wood and stone and various materials, which appear over the wall? Some of them are talking, others silent.
>
> You have shown me a strange image, and they are strange prisoners.
>
> Like ourselves, I replied; and they see only their own shadows, or the shadows of one another, which the fire throws on the opposite wall of the cave. (Jowett, 1999: 209).

As the prisoners converse, and without any other vantage point, they naturally perceive the shadows to be reality – they assign names and try to explain the various flickering shapes which appear on the wall. Echoes of sound are attributed to the shadows. Consider what happens if a prisoner were to become unchained, and look into the light. Once his eyes adjusted, he would realize that "what he saw before was an illusion (Jowett, 1999: 210)." However, any efforts to explain the true nature of reality would be immediately ridiculed by his peers.

What relevance does Socrates' allegory have for strategic management researchers? We propose that the cave is a metaphor for one of the most serious threats to strategic management research: poor construct measurement. While the implications of measurement error are well known, they are typically ignored in a majority of studies on strategic management topics. So, as with the freed prisoner, many academic researchers ignore or are unaware that their measures often may not fully or accurately capture the constructs of interest. Our purpose is to highlight both the extent and consequences of measurement error in strategic management research.

We begin with a brief overview of research design and methodology issues in strategic management. Next, we explore several topics in more detail, including statistical power, sample size, and measurement. To illustrate the consequences of measurement issues, we replicate a prominent debate among strategy researchers: whether diversification is a consequence of agency costs (Amihud & Lev, 1981). Using data from 640 *Fortune* firms, we created multiple indicator models of both agency costs and diversification. Then, we developed multiple LISREL models representing all possible single- and multiple-indicator combinations of these variables. Our results provide strong evidence that the debate between authors (i.e., Amihud & Lev, 1999; Denis, Denis & Sarin, 1997; 1999; Lane Cannella & Lubatkin, 1998; 1999) is largely an artifact of measurement error. Finally, we examine the potential for similar problems in other strategic management areas. We assess the "state of the art" of measurement in strategic management research with a review and critique of 196 empirical strategic management articles published in a recent three-year period.

RESEARCH DESIGN ISSUES IN STRATEGIC MANAGEMENT

Background

Strategic management is generally acknowledged to be one of the younger subdisciplines within the broader management discipline. Such emergent areas are typically characterized by debate, and challenges to existing paradigms (Kuhn, 1996). While the latter are often couched as theoretical discussions, empirical work plays a critical role in confirming, or disconfirming, a particular perspective. Contributing to this advancement of the field, there has been a small research stream that critiques empirical research in strategic management. This stream includes both narrative (Hitt, Gimeno & Hoskisson, 1998; Venkatraman & Grant, 1986) and quantitative reviews. Examples of the latter are summarized in Table 1. Regardless of the topic, these reviews have been consistently critical of the rigor of strategic management research. However, one critical dimension of research design – construct measurement – is not covered by this pool of studies.

Insert Table 1 around here

Construct measurement is particularly relevant to strategic management research, as the variables of interest tend to be complex or unobservable (Godfrey & Hill, 1995). Paradoxically, measurement has historically been a low priority topic for strategic management scholars (Hitt, et al., 1998). As a result, complex constructs have often been represented with simple measures, and with limited testing for reliability or validity (Venkatraman & Grant, 1986). Our intent is to contribute to this research stream with a critique of measurement issues in strategic management. We begin with a brief discussion of two related topics – statistical power and sample size, and the compounding effects of measurement error.

Statistical power and sample size

Power is the potential of a statistical test to yield a significant result. Issues relevant to power include a researcher’s willingness to consider Type I and Type II errors, sample size, the magnitude of the effect being examined, the test being used, and the quality of the data (Cohen, 1988, 1992). **Type I** error is the risk of mistakenly rejecting the null hypothesis – i.e., a false positive finding. This type of error is routinely addressed in empirical studies, with the $p \leq .05$ level (alternately stated, at least a 95 percent likelihood that a relationship is not a false-positive) widely accepted as an appropriate threshold. **Type II** error is the risk of failing to reject the null, even though a meaningful relationship does exist. Statistical power (p) is an estimate of the probability that a null hypothesis will be rejected for a given effect size. Cohen (1988) recommends using 0.80 as the threshold for power assessment – i.e., an 8 in 10 chance that an existing relationship will be successfully detected.

Everything else constant, a more stringent p-level for a Type I error leads to a greater risk of Type II error, and vice versa. Additionally, sample size affects the risks of Type I and Type II error. For example, consider a population correlation between two constructs of 0.30. Using a criterion for statistical significance of a=.05, a sample of N=30 has only a 50 percent chance of successfully detecting this relationship. The probability improves to 70 percent with the addition of 20 subjects, though, and over 90 percent when the sample is increased to 100 subjects.

While the Type I error is considered by authors and reviewers alike, Type II error frequently is overlooked; surveys of management authors reveal that power analyses are unusual, and that the perceived need for such analysis is low (Mone, Mueller & Mauland, 1996). In stark contrast to these surveys, the power of most studies is weak: strategic management studies have been characterized as having only *half* the recommended power levels; that is, they have only a 4 in 10 chance of rejecting the null hypothesis (Mazen, Magid, Hemmasi, & Lewis, 1997). More recently, Ferguson and Ketchen (1999) reviewed the research stream on organizational configurations, and concluded that only 8 percent of published studies had sufficient statistical power. Finally, Mone and colleagues (1996) reported that the statistical power of many strategic management studies was significantly lower than in several other management subdisciplines.

Thus, there is strong evidence to conclude that statistical power is a critical issue in the design of academic research, and that the power of studies in strategic management has been weaker than others. In the next section, we explore the status of construct measurement in strategic management, another potential problem, and the implications of measurement error for statistical power.

Measurement error and attenuation

Blalock (1979) described models of social processes as having three elements: (1) a theoretical language that describes causal relations between constructs; (2) an operational language that links certain indicators to their respective constructs; and (3) an integrative theory that links the causal ties between constructs and indicators. The second component is of particular relevance to strategy research.

Most strategic management research is framed using Blalock's first component – a hypothesis that relates two unobserved concepts. So, for example, research may posit that the presence of an agency problem can lead to opportunistic actions by executives. However, this hypothesis is not tested directly. Instead, a researcher may study the relationship between two variables (or indicators) which serve as proxies for the respective constructs: For instance, CEO equity or the ratio of insiders on the board may be used to predict levels of executive pay; CEO equity serving as the proxy for agency problems, and pay serving as the proxy for opportunistic behavior.

If the indicators fully represent the latent concepts, power is unchanged. In practical terms, this requires that all variables are valid, and measured without error. However, even moderate amounts of measurement error can have substantial negative implications for power

(Schmidt, Hunter & Urry, 1976; Zimmerman & Williams, 1986). Power analyses do not consider measurement error – instead, the calculations to determine a minimum *N* assume exact measurement of predictor and outcome variables. Consequently, even researchers who conduct power analyses "will take samples that are too small and will be too unlikely to reject the null hypothesis, even when a reasonable hypothesis is actually true" (Maxwell, 1980: 253).

As an example, consider a researcher who is designing a study and the population effect size is believed to be moderate to small (r=0.30). Setting p=.05, a sample size of 150 is needed to have a power level of 0.80 (Cohen, 1988). However, Cronbach's alpha for predictor and outcome variables are each 0.60. Because of the measurement error associated with each term, the observed correlation will be much smaller – approximately r = 0.10. The sample of 150 now has only a 1 in 3 chance of detecting the observed relationship.

As we have described previously, statistical power levels are often unacceptably low, in both the broader management field, and in strategic management research in particular. (Mone et al., 1996; Ferguson & Ketchen, 1999). More importantly, the presence of measurement error indicates that prior reviews may actually *underestimate* the magnitude of this problem: "The bottom line is that unreliability shrinks observed effect sizes and therefore reduces power, and increases in reliability enhance observed effect sizes and therefore increase power" (Cohen, 1988: 537). In the following section, we present two analyses that further describe the potential severity of this problem. In the first, we describe how measurement errors can mistakenly fuel divergent findings and perspectives. The second analysis examines the extent of measurement error problems in strategic management research.

Study 1: Illustrating the Consequences of Measurement

Our review of research design issues indicates that measurement practices may negatively affect the rigor and generalizability of strategy research. Because this research serves as the basis for new theory development along with normative guidelines, textbooks, and practitioner articles, it is essential that we determine whether measurement constitutes a small or large threat to the validity of the research base. To explore this issue, we examined how varying levels of measurement affect the outcome of a prominent hypothesis: Amihud and Lev's (1981) agency model of diversification. We begin with a review of relevant research, followed by a discussion of the relevance of this stream to the concern for construct measurement. We then demonstrate the consequences of measurement problems using a series of structural models.

Background

A common explanation for diversification is the continued search for growth. A mature firm might consider expanding the scope of its offerings in pursuit of new growth opportunities. An alternate explanation is based in agency theory. Much as investors strive to balance their personal portfolios and thus their risk, agency theorists contend that top managers expand the firm's business portfolio to mitigate their individual risk -- even if doing so ultimately results in a reduction of shareholder wealth.

Evidence suggests that the unique interests of managers, including natural inclinations toward risk aversion (Berle & Means, 1932; Jensen & Meckling, 1976), help to explain many organizational phenomena including executive perquisites (e.g., Jensen and Murphy, 1990; Lambert, Larcker, and Weigelt, 1993), governance innovations (e.g., Singh and Harianto, 1989), and strategic initiatives (e.g., Baysinger, Kosnik, and Turk, 1991; Sirower, 1994), among others.

The latter explanation has achieved the status of conventional wisdom in the two decades since Amihud and Lev's (1981) examination of the relationship between corporate ownership structure and diversification strategies in 309 large U.S. firms. Their study revealed that management-controlled firms engaged in conglomerate mergers at a far greater rate than owner-controlled organizations. Because conglomerates are typically valued at a discount – much to the disadvantage of shareholders (Berger & Ofek, 1995; Denis, Denis, & Sarin, 1997), Amihud and Lev (1981) concluded that managerial self-interest is a primary motivator behind diversification.

Relevance

Three factors guided our selection of Amihud and Lev's work to illustrate the consequences of measurement error: First, while their results have been largely accepted in the field, their work has recently been challenged. Second, there are issues surrounding the measurement of both predictor and dependent variables. Third, statistical power and attenuation play a role in interpreting results to date. Next, we discuss each of these issues in more detail.

Challenges to conventional wisdom. As noted in our introduction, debate and challenges to conventional wisdom are central to a field's advancement (Kuhn, 1996). Recently, Lane, Cannella, and Lubatkin (1998) challenged the validity of the agency model of diversification, Amihud and Lev's (1981) results. Lane and colleagues attempted to replicate Amihud and Lev's (1981) research and also updated the examination of diversification activity to a 1980s timeframe. They concluded that owner monitoring – or control – had little effect on corporate diversification strategies. The debate between these researchers was highlighted in a recent issue of *SMJ*. In the same volume, Denis, Denis, and Sarin (1999) summarized the matter, noting that:

> "Though both sets of authors conduct similar empirical tests on virtually identical data, they arrive at completely different conclusions. Lane et al. (1999: 1077) conclude that '…there is little theoretical or empirical basis for believing that monitoring by a firm's principals influences its diversification strategy and investment decisions.' In contrast, Amihud and Lev (1999: 1064) conclude that 'The evidence shows that there exists a relationship between corporate diversification and corporate ownership structure.'" (page 1071)

Measurement issues. Denis and colleagues (1999) argued that resolution of this debate hinges, in part, on a careful evaluation of the empirical evidence. Their own review suggested that the methodologies of both studies were flawed, with an important shortfall noted in the studies' measurement approaches. For example, each used broad ownership categories constituting coarse-grained indicators of agency conditions (e.g., McEachern, 1975; Palmer, 1973). When improved constructs were substituted in the analyses – namely, ratio-level indicators of equity ownership, as well as more refined measures of diversification – more substantial results were generated (Denis, et al., 1997; 1999).

Echoing Denis, et al. (1999), we believe that the confusion surrounding the agency – diversification link is, to a significant extent, an artifact of methodology, and most especially, measurement. Empirical analysis confirms that measurement error is more prevalent for abstract versus concrete concepts (Cote & Buckley, 1987). Since the publication of Amihud and Lev's (1981) work, the field's understanding of the key variables – most notably, board control and diversification strategies – has advanced considerably. So, too, has our ability to measure these variables. In the context of control alone, it is now well recognized that the construct has several nuances (Fama & Jensen, 1983), leading researchers to recommend use of multiple measures when studying control issues (Eisenhardt, 1989). Recognizing the complexity of measuring board oversight, one study developed a multi-indicator factor model to tap control (Boyd, 1994).

There are similar opportunities to refine the measurement of firm diversification. While there are multiple measurement schemes available – including Rumelt's categories and SIC counts – the entropy measure (Palepu, 1985) has been reported to have superior reliability and validity (Chatterjee & Blocher, 1992; Hoskisson, Hitt, Johnson & Moesel, 1993). The entropy measure is particularly germane to our analysis, as it can be decomposed into unique elements – indicators of both related and unrelated diversification (Acar & Sankaran, 1999; Palepu, 1985).

Power. Of the core studies in this research stream, only Lane and colleagues (1998) have explicitly addressed statistical power. They argued (1998: 563) that their sample size of 309 had ample power, as "Cohen (1992: 13) observed that economic research usually reports large effect sizes." Additionally, they also noted that their sample had ample power to detect moderate effect sizes as well. However, Cohen (1988) stated that the expectation of large effect sizes may hold only when using "potent" variables, and/or in the presence of strong experimental controls. Separately, Cohen (1988: 413) also commented that, "what may be a moderate theoretical effect size may easily, in a "noisy" research, be no larger than what is defined here as small."

Sample size requirements change dramatically, depending on expected magnitude of the effect being studied. Cohen (1992: 158) offered a simple comparison: Consider a regression model with three predictors; the researcher desires a significance level of p =.05, and an 80 percent likelihood of successfully detecting the relationship. Minimum sample size is 34 for a large effect, 76 for a moderate one, and 547 for a small effect. Lane et al. (1998) sampled 309 firms, and Denis et al. (1999) sampled 933 firms. Therefore, if there is a moderate theoretical effect size between agency factors and diversification, and measurement error exists, only Denis et al. likely had sufficient power to capture an attenuated effect.

The purpose of our study is to refine the debate surrounding the control-diversification relationship. We build on the methodological refinements recommended by Denis, et al. (1997; 1999) and other scholars (e.g., Boyd, 1994; Eisenhardt, 1989) to test a series of models that use progressively more fine-grained measures of both variables – corporate control and extent of diversification. These measures are not only consistent with recent advances in strategic management research, but should provide a more robust test of the agency relationship. Stated formally:

H1: Board control is negatively related to the level of diversification.

H2: The relationship between board control and diversification is stronger when both variables are measured with multiple indicators.

Methodology

Sample. Data were collected from a random sample of 640 *Fortune* firms as part of a larger research project. The sample included over 50 2-digit SICs, and nearly 200 4-digit SICs. Company names were selected randomly, and proxy statements were used to collect governance data. Our design is cross-sectional, with all data from the year 1987.

Analysis. In order to test for the effects of measurement error and attenuation, we tested our hypotheses in a structural model, using LISREL VII. Consistent with the approach taken by Denis et al. (1997), we used the extent of diversification as the dependent variable, versus merger activity. The model is shown in Figure 1.

Insert Figure 1 around here

Measurement. **Board control** was measured using Boyd's (1994) multi-indicator factor model[1]. The indicators for this measure are duality (CEO and chairman), ratio of insiders, director stock ownership, representation by ownership groups, and director pay. Proxy statements were used to code these variables. Duality and director pay load negatively on this construct, while the other indicators load positively. **Total diversification** (Palepu, 1985) was separated into its components *du* (unrelated) and *dr* (related), using data from the Compustat Business Segment database and company 10-K filings. Finally, we included **firm size** as a control variable, as it has been previously linked to levels of diversification (Denis et al., 1997). We measured size with three indicators: net sales, total assets, and total stockholder equity, also from Compustat. Log transformations were used to normalize all size indicators.

Results

Descriptive statistics for all variables are reported in Table 2.

Insert Table 2 around here

Tests of dimensionality. Prior to testing hypotheses, we conducted a series of analyses to confirm the factor loadings and dimensionality of our predictor and control variables. The first model was a confirmatory factor analysis for the board control construct. The results of this analysis are consistent with Boyd's (1994) results. All factor loadings were in the expected direction, and statistically significant at the p=.001 level or greater. Overall fit measures reported that a unidimensional model was best fit to the data.

[1] As noted by our reviewers, Boyd's model is not an exhaustive set of agency indicators. Consequently, we conducted additional analyses to evaluate the robustness of our results. We developed new models that introduced a sixth indicator, CEO tenure, as another measure of board oversight. While tenure loaded significantly on the board control factor model, its magnitude and significance were substantially less than the other extant indicators.

Second, we examined whether it is appropriate to treat *dr* and *du* as indicators of a common dimension. The full model presented in Figure 1 provides strong support for this assumption: *dr* was used as the referent indicator (a loading of 1.0), and the loading for *du* was 0.63, statistically significant at the p=.01 level.[2] However, an alternate argument could be made that the related and unrelated diversification strategies are different phenomena, and as such likely have differing relationships with agency variables. For instance, managers might consider related and unrelated portfolios to have different types and levels of risk. If true, *dr* and *du* would have unique associations with ownership or monitoring variables. While such differences would be reflected in the Figure 1 factor loadings, we conducted an additional test for purposes of rigor. We tested this competing perspective in a supplementary model that treated *dr* and *du* as independent constructs, and having separate paths from control and firm size – i.e., a seemingly unrelated regression. Using an incremental chi-square test, this alternate model had a significantly worse fit than the Figure 1 model. These results, coupled with the strong factor loading for *du* in the hypothesized model, provide strong support for our multi-indicator approach. Finally, factor loadings for the three size indicators were highly significant and in the expected direction.

Model summary statistics. Coefficients were statistically significant and in the expected direction for all structural and measurement paths in Figure 1. Overall model measures reported a very good fit: Goodness of fit was 0.94; the root mean square residual was 0.08; other measures reported comparable fit. The coefficient of determination, or R^2, was 0.25 for the dependent variable(s). In comparison, we explain 50 percent *more* variation of this variable than

---

Therefore, inclusion of a sixth indicator yielded only minor changes in path coefficients, and the results of hypothesis tests were unaffected.

Denis and colleagues' (1997) analyses do, despite using five *fewer* control variables. There was a statistically significant, negative covariation between control and firm size (phi = -.28, p=.001); in other words, governance oversight tended to be weaker in larger firms. Firm size has a positive effect (0.11, p=.01) on diversification as well.

Hypothesis tests. **Hypothesis 1** was supported, with a statistically significant, negative relationship (γ = -.15, p=.05)[3] between board control and diversification. **Hypothesis 2** stated that the relationship between control and diversification would be stronger when using multiple versus single indicators. To test this hypothesis, we constructed twelve additional LISREL models; these are reported as sets 2, 3(a) and 3(b). Summary statistics for these models are reported in Table 3, and provide strong support for hypothesis 2.

Insert Table 3 around here

In set 2, we leave the confirmatory factor model for control unchanged, but treat diversification as a single-indicator construct. This yields two separate models – one with *du* representing diversification, and a second using *dr*. By comparing these models against the model in Figure 1, we can identify the degree of attenuation associated with less precise measures. In this set of models, the magnitude of the path coefficient decreases slightly (from -.15 to -.14, on average); however, the explained variance of the overall model (retaining control for firm size), drops from 0.226 to 0.029, on average. Additionally, the path coefficient for

[2] Because there are only two indicators for this dimension, it is not feasible to conduct a separate confirmatory factor analysis for diversification.
[3] We tested for the possibility of reverse causation – i.e., that levels of board control are a *result* of the firm's diversification posture – in a supplemental model that treated diversification as a predictor and control as the dependent variable. The path between these two variables was not statistically significant, and overall levels of explained variance were smaller than our hypothesized Figure 1 model.

control becomes nonsignificant when *dr* is used as the sole indicator. In other words, the use of single indicators for the dependent variable is associated with (a) much greater likelihood of Type II error, and (b) in the case of statistically significant results, substantial underestimation of the magnitude of the relationship. To maintain consistency with the simplified measurement schemes of these sub-models, all of the set 2 and set 3 analyses used net sales as the sole indicator of firm size.

In set 3, both board control and diversification were measured with single indicators – i.e., we introduce measurement error on both sides of the equation. These included five models based on *du* and each of the control measures (subset 3a), and another five models based on *dr* and each of the board control indicator measures (subset 3b). Director pay was linked with both *dr* and *du*, while duality was linked only with *du*. The strong results for director pay is not surprising, as it had the strongest factor loading in Boyd's (1994) analysis. The positive links for these two predictors are expected, because levels of director pay and the presence of duality are associated with increased likelihood of agency problems – i.e., these variables load *negatively* on the control construct.

The models of set 3 also supported hypothesis 2. When single indicators were used for both the predictor and outcome variables, seven of the ten hypothesis tests were not statistically significant. Additionally, the path coefficient changed directionality in half of the models. The corresponding reduction in explained variance was substantial: In comparison to the full model (CED = .25), the aggregate explained variance was 0.020 for subset 3a, and 0.025 for subset 3b. In other words, there was a 10X magnitude of difference in explained variance between the full model and single indicator models. Additionally, in seven of the ten tests, the single indicator models would lead to a mistaken conclusion that there was no relationship. These results provide strong support for hypothesis 2.

Our results suggest that there is an element of truth in both the Amihud and Lev and Lane et al perspectives – while there is a significant linkage between agency conditions and diversification, it is limited in magnitude. More importantly, however, this debate appears to be largely an artifact of measurement problems. Previously, Denis and colleagues (1999) argued that resolution of this disagreement hinged on a careful review of methodological issues. As we have demonstrated here, discordant results are highly likely when the relevant variables are measured with single indicators. As we demonstrate with the set of single-indicator models, there is a 70 percent probability of mistakenly concluding there was no relationship between control and diversification. Equally important, such research underestimates the true effect size by a magnitude of almost 10.

Additionally, by decomposing the elements of diversification, we find that there is a differential effect for the type of diversification: High levels of board control have a stronger effect on unrelated versus related diversification. In other words, while firms with strong board oversight are less likely to have any type of diversification activity, the effect is more pronounced for unrelated as opposed to related diversification. These results are demonstrated in the factor loadings for our full model, as well as the varying results for set 2 and set 3 submodels. This differential may reflect an assessment by the board that related diversification has greater potential value to the firm.[4]

[4] This point was highlighted by one of our reviewers.

Content Analysis of Empirical Studies

We've highlighted the effect of approaches to measurement and measurement error in one context – the debate between ownership and diversification – but to what extent is measurement error an issue in the broader pool of strategic management research? To explore this question, we completed a content analysis of a sample of strategy-related articles published in leading scholarly journals.

Sample

To determine the optimal characteristics of the sample, we reviewed the design characteristics of prior methodological critiques of strategic management research, as shown in Table 1. This review suggests that the sample should have three attributes: First, a range of journals should be sampled. Second, the sample should have a multi-year time frame. Third, prior critiques have used two different approaches for selecting specific articles for analysis – some have included all articles that met the relevant criteria (e.g., Ketchen & Shook, 1996; Mone, et al., 1996), while others included a subset of relevant articles (e.g., Hubbard, et al., 1998). We chose to include all relevant studies in the interests of generalizability.

Our sample comprised the universe of empirical strategic management articles that were published in the discipline's leading scholarly outlets. We began with MacMillan's (1989) set of fourteen primary outlets for strategy research wherein a half dozen of these journals were ranked by an expert panel to be of 'outstanding quality': *ASQ*, *AMJ*, *AMR*, *HBR*, *MS*, and *SMJ*. We excluded *AMR* and *HBR* from our list as they generally do not publish empirical work. Therefore, strategic management articles published in the *Academy of Management Journal*, *Administrative Science Quarterly*, *Management Science*, and *Strategic Management Journal* were included in the final sample. We selected a time period of 1998-2000, as the recent studies are presumably most likely to have the highest level of methodological sophistication. Collectively, our combination of leading journals and recent timeframe should provide a “best case” assessment of the state of construct measurement in strategy.

We reviewed each volume of the journals selecting articles for inclusion using a two-stage approach. First, we identified articles reporting research on strategic management topics. We included all papers published in *SMJ*, as it is a discipline-specific outlet. From *AMJ*, *ASQ*, and *MS* we selected all articles that met a liberal definition of falling into the strategic management domain. The coders who made these assignments have served as reviewers on manuscripts and have held repeat editorial board assignments on a subset of these journals. As our focus was on the *use* of measurement approaches, as opposed to the development of such approaches, we narrowed the pool by selecting only articles reporting empirical tests of hypotheses. We specifically excluded those relying solely on case analysis and descriptive statistics, those using meta-analyses as they are restricted in their selection of measures, and those developing measures but not testing hypotheses. For example, if an article included only the development of a scale, it was excluded as the *use* of the scale in a test of hypotheses was not an aspect of the study as an independent statistical test. If an article developed and validated a scale and also *used* the scale in a hypothesis test, the article was coded as having one statistical test and is included in our sample. This screen yielded a final sample of 196 articles – a sample comparable to the prior methodological critiques listed in Table 1. A list of the sample articles is available from the authors. As a post hoc analysis, two external raters unaffiliated with the project independently assessed a random sample of 70 articles from *AMJ*, *ASQ*, and *MS*. Agreement with inclusion and exclusion decisions was consistent with our ratings in this study (alpha = .91).

A substantial number of the articles included multiple statistical tests with different independent and dependent variables, samples, sample sizes, and analyses. Therefore, each *statistical test* was used as the unit of analysis. We selected the most complete models presented as multiple, hierarchical-like models were common (e.g., a regression with control variables, indicators, and interaction terms tested in three separate models). We counted all tests that utilized different dependent variables and samples as unique. To avoid allocating extra weight to articles that present multiple sub-sample analyses, we counted sub-samples only if a new dependent variable was utilized. This yielded a final sample of 625 statistical tests from the 196 articles – a sample considerably larger and more comprehensive than prior methodological critiques in the field.

Analysis

A content analysis of each article and test was completed by an expert rater. A subset of articles was coded by a second rater with comparable results (alpha = .96). The articles were examined to evaluate the construct operationalizations employed with the intent of developing a categorization scheme. We elected to treat all variables as potential constructs. This was based on two findings from the review: differences between articles regarding what constitutes a construct and the prevalence of "hidden" constructs masked as single variables within the studies.

We found that the definition of what constituted a construct within the strategic management literature was largely at the discretion of the researcher. Constructs are "theoretical creations based on observations but which cannot be observed directly or indirectly" (Babbie, 1989: 109) and the basis for most strategic management theories (Godfrey & Hill, 1995). In practice, we found many constructs in the sample were not identified as such. Organizational size, for example, appeared in our sample as a widely utilized variable and is arguably one of the most commonly used variables in the strategic management research. Size was repeatedly found in the samples as an independent and a control variable. We drew on a sub-sample of articles and found “size” as a proxy of available organizational resources, propensity/ability to initiate competitive action, core rigidity, and public profile among a wide range of other constructs. These varied constructs were, however, generally operationalized using a single indicator. No attempts to examine (establish) convergent validity were reported for the association between size and other measures of the intended construct. As single indicators were the norm, reports of reliabilities and measurement error were not common.

When examined across the volume of studies, the measures of size also appeared to vary far less than the constructs size was purported to represent. Three indicators in particular – total assets, sales, and employees –constituted over 80 percent of the size measures employed despite the range of constructs size presumably represents.

Because of the lack of specific criteria for identifying constructs in the studies and the potential for commonly used variables to represent complex constructs, we decided to treat all variables as potential constructs. The primary benefit of this approach is a lack of positive bias in the use of multiple indicators in the field for constructs. However, our analysis should be considered as a comprehensive yet conservative estimation of the use of construct measurement.

The measures employed in the tests were coded into one of five categories that progressively provide increased ability to assess validity, reliability, and measurement error. These categories are: single indicators, discrete items, single ratios, indexes, and scales / multiple measures. S*ingle indicators*, at the nadir of methodological sophistication, provide the researcher with the least assurance that a measure is a valid and reliable proxy of a construct and no estimates of reliability, and thus error, are possible. In the context of a regression model, we coded the use of a sole variable with an accompanying beta as a single variable. For example, a regression estimate for "sales outside of home country" was coded as a single indicator if no other variables for internationalization were included.

Single ratios, like single indicators, serve as sole indicators of a construct but they are comprised of two parts in the form of a ratio. These variables may provide an advantage over single indicators as they allow a multi-faceted perspective (i.e., condition *Y* in relation to *Z*) but they may also mask important information and do not allow for the overall association between the variables to be examined in terms of reliability – a limitation identified by one of our reviewers. For example, a single ratio of the internationalization construct is "ratio of foreign sales to total sales". Other common single ratios include debt-to-equity, return-on-assets, and book-to-market measures (e.g., Tobin's *Q*).

Discrete indicators are collections of single indicators that collectively serve to indicate a construct. They are conceptually linked, but have their own beta estimates in a regression model. For example, the internationalization construct may be assessed using three separate, discrete variables (e.g., "sales outside of home country", "employees outside of home country", and "count of products sold outside of home country"). The correlation between discrete items can be analyzed and serve as a limited assessment of reliability.[5]

A second category of measures, all forms of multiple measures, allows formal assessment of reliability and thus error to be quantified. Two such approaches are indexes and scales. *Indexes* incorporate measures of one or more dimensions of a construct into a single item,

[5] It is important to note indicator causality here. In most applications, indicators are seen as effects of an underlying construct. Other times, however, the indicators may drive the construct. In this context, described as causal or formative indicators, diagnostic tools such as inter-item correlations or reliabilities may not be relevant (MacCallum

commonly using a summative approach. For example, a firm's level of internationalization can be operationalized using an index calculated by summing "foreign sales", "foreign employees", and "expatriate managers". In a regression model, a single beta is calculated for the index. Indexes commonly utilize scale-dependent weights of each indicator that comprise the index or category weights assigned from a distribution of sampled subjects. The reliability among index components may be calculated prior to index creation to provide statistical support for the collective measure.

The final category, s*cales and multiple measures*, utilize data reduction approaches (i.e., factor analysis, principal component modeling, structural equation modeling, etc.) to explicitly assess the degree to which multiple items represent a construct and the error associated with the measure. In a regression model, a single beta is calculated for the scale, not for the individual items. For the internationalization construct, a single "internationalization" value may be comprised using multiple items (e.g., "foreign sales", "foreign employees", and "expatriate managers") and assessing fit onto a common dimension. Assessing reliability, and thus measurement error, is inherent to such an approach.

Results

The presentation of the results of our content analysis is focused on three areas: sample size, measurement schemes, and reliability. Highlights of these results are shown in Tables 4 and 5.

Insert Tables 4 & 5 around here

& Browne, 1993). From an anecdotal review of our article pool, the notion of indicator causality is discussed only occasionally. We would like to thank one of our reviewers for this important point.

Sample size. Other studies (Mazen, et al., 1997; Mone, et al., 1996) have documented the scope of power issues in strategy research. Consequently, a power analysis of our sample would add significant length to the paper, yet little new knowledge**.** Two aspects relating to statistical power, sample size and the ratio of sample size to indicators, are germane to our analysis, however.

We begin with an examination of sample size statistics. In aggregate, the mean sample size[6] of our pool was 2,559 (*s.d.* = 12,909), and ranged from 20 to 158,782. The mean alone is misleading given a high positive skew (S = 8.42) and large kurtosis (K = 83.41) both of which are highly significant. A more accurate picture of the average sample size is garnered from an examination of the percentiles. Studies at the 50th percentile utilized a sample of N = 215 with studies at the 75th, 90th, and 95th percentiles having samples of 426, 1,461 and 8,699 respectively.

The number of indicators used in a study relative to the sample size also provides an indication of statistical power. The distributions for the number of independent variables and control variables along with the ratio of sample size to independent, control, and the sum of independent and control variables are significantly skewed and non-normal. A statistical test at the 50th percentile was comprised of 4 independent variables and 3 control variables having a sample size to independent variable ratio of 58 to 1, a sample size to control variable ratio of 39 to 1, and an overall ratio of 24 observations to independent and dependent variables. Two conclusions may be drawn from this analysis. First, while ultimately dependent on the statistical power of the measures used, ratios of subjects to indicators in the studies examined appear consistent with generally accepted norms. Second, and directly related to our emphasis on measurement, sample size insufficiencies do not appear to be a hindrance for the use of

---

[6] In assessing sample sizes we did not adjust for the effective reduction associated with the use of pooled cross-sectional data

measurement schemes incorporating multiple measures. The sample sizes, on average, appear to allow incorporation of multiple measurement approaches into the research designs without the burden of obtaining additional subjects.

Measurement schemes. At the nadir, the use of measures which cannot be assessed for reliability (i.e., single indicators, single ratios, and discrete items) provide the researcher and reader with the least assurance that a measure is a valid and error-less proxy of a construct. Our review of the published tests suggests that the use of such measures is a common, but not exclusive, approach in strategy research.

Results indicate that much of the published research does not pay attention to the problem of measurement error. Fully 70.8 percent of independent, 57.7 percent of dependent, and 92.7 percent of control variables are based on a methodology that forgoes the assessment of reliability.

Of the 3,388 independent variables, 1,613 (47.6 percent) were single indicators. Of the 677 dependent variables, 228 (38.1 percent) were measured using single indicators. For control variables, 79.6 percent (4,280 of 5,376 variables) were single indicators. Across the 625 tests reviewed, 335 used at least one single indicator for an independent variable, 238 for a dependent variable, and 371 for a control variable. Thirty-four (5.4) percent of the sample relied on single-item indicators *exclusively* for independent, dependent, and control variables. Single ratios also appear prevalently in the studies examined comprising the measures for 9.1 percent, 14.9 percent, and 8.1 percent of the independent, dependent, and control variables. Discrete items comprise the measures for 14.1 percent, 4.7 percent, and 5.0 percent of the respective variables.

Techniques allowing reliability and measurement error to be assessed are used in some of the strategic management studies and should be noted. A small but laudable number of tests, only .32 percent of all examined in our study, utilized a full compliment of multiple measures. These tests operationalized their measurement using either indexes or scales for *all* of the IV's, DV's, and control variables. Unfortunately, such rigor was the exception rather than the standard practice.

Indexes and scales / multiple measures comprise 14.8 percent and 11.8 percent of the independent variables used respectively, 17.3 percent and 19.6 percent of the dependent variables, and 1.9 percent and 0.7 percent of the control variables used. Additionally, 17.4 percent of the studies rely solely on these types for independent variables and 25.1 percent for dependent variables. However, less than one half of one percent of the studies rely on control variables which allow reliability and measurement error to be assessed. Control variables serve the purpose of ensuring that the predictions provided by independent variables under examination are not overly inflated due to covariance with variables suggesting other explanations. If measures of control variables are not reliable proxies for their intended constructs, and our analyses suggests there is little evidence supporting reliable measurement, the true value of the explanatory variables is likely inflated.

Reliability. While measures that allow reliability to be addressed are desirable, the reporting of reliability information is necessary to obtain the full value from this effort. Such information can be provided explicitly or implicitly in the manuscript. We find that reporting of this vital information is not universal practice.

The multiple measure approaches utilized within the research appear sound. The reported reliabilities in general are acceptable (average α = .80, .82, and .76 for scales / multiple measures used as independent, dependent, and control variables respectively). These outcomes may, however, be an artifact of the reporting as the majority of studies do not report reliabilities.

Reliability assessments were reported for only 0.5 percent of the independent and dependent variables based on indexes. Surprisingly, control variables based on indexes receive the most attention - nearly half (44 percent) of the studies using multiple indicators for control variables report reliabilities. This result indicates that the validity of the indexes is generally assumed and not subjected to statistical confirmation

While the reporting of reliabilities for indexes is uncommon in the studies we reviewed, studies using scales did somewhat better. Of the 399 independent variables based on scales, reliability scores are reported for only 133 (33.3 percent). The reporting rate for the reliability of dependent variables measures was higher, 63.2 percent, while the rate for measures of control variables was approximately 41 percent.

While less informative than reliabilities, correlations between indicators provide an indication of reliability. Correlation matrices were presented in fully 142 of the 196 articles (71.4 percent). Approximately one third of these (49 articles; 25 percent of the total) included all variables in the correlation matrix. This is a positive finding, but we also regularly found independent variables, dependent variables, interaction terms, squared terms, and control variables missing from the correlation matrixes. Most disturbing, fully one quarter of all articles, 50 of the 196, present no correlation matrix.

As a follow-up to Mone et al.'s (1996) call for greater attention on statistical power in strategy research, we examined the studies for evidence of attention to statistical power. Of the few studies that present a power analysis, none incorporated the reliability of their measures into their calculation. An understanding and appreciation of the influence of poor construct measurement on both the *a priori* planning of research approaches and the *post hoc* diagnostics of results appear to be absent

The Typical Article

Thus far we have focused largely on the statistical tests in reporting results. An analysis of the measurement approaches used in a typical article may place these findings in a context more familiar to researchers.

Measurement Approaches Used The typical strategic management article examined in our study includes 17 independent variables and 27 control variables **(**We operationalized 'typical' based on the averages per 196 articles examined). While regression was the most common statistical tool, some studies utilized structural modeling, path analysis and similar approaches resulting in an average of 3.5 dependent variables across the 3 statistical tests in the typical article. The independent variables utilized in an article, on average, consisted of 8.3 single items, 1.6 ratios, 2.4 discrete measures, 2.6 indexes, and 2 scales. This is an encouraging result as the typical article published in leading strategy journals utilizes almost five independent variables (roughly thirty percent) for which reliability can be assessed. These results are shown in Table 6 and Figure 2.

Insert Table 6 & Figure 2 around here

For dependent variables, the results are comparable. On average, the articles in our sample utilize approximately 2 dependent variables for which reliability could not be assessed and only one that could. However, less than half of the studies that could report reliabilities actually did so. For control variables, the typical study relies virtually exclusively on single item approaches. Fully 97 percent of the average 27 control variables used in the average article used single items, ratios, or discrete measurement approaches.

A content analysis of a subset of articles indicated that the majority of these articles reference prior research as support for general measurement approaches.[7] Over half of the total variables were supported with references to prior work. The references ranged from very general support for expected relationships between variables of interest (e.g., correlations, results of hypothesis tests, etc.) to references for specific variable measures.

One quarter of the independent variables examined in the subset included references to a specific previously published source for a particular measurement approach. Approximately one third of this support was explained in terms such as "based on" or "a modified version" of a published measure. For dependent variables, the use of established measures was even more common. In fully two-thirds of the dependent variables prior work was cited to support a specific operationalization. Referencing for measurement approaches that allow reliability to be assessed and for approaches that do not was roughly equally divided. Using references to prior work for support was less common for control variables with cites for only ten percent of the measures used.

Thus, poor measurement within these articles, in many cases, may be the result of a reliance on previously published work regardless of the quality of that measurement approach. This conclusion suggests that reviewers and journal editors share some of the responsibility, along with authors, for the persistence and pervasive reliance on poor measures.

DISCUSSION AND CONCLUSIONS

As an academic specialty, strategic management is a relatively young discipline: Depending on the metric used, the field is between two and three decades old. However, even

---

[7] We'd like to thank one of our reviewers for recommending this approach

with this youth, it plays a critical role in the study of business and management. As the field has matured, there are increasing expectations for the rigor of strategic management research. Our purpose is to extend the ongoing commentary on methodological issues by highlighting the importance of construct measurement. As demonstrated by the content analysis presented herein, there has been little emphasis placed on measurement concerns in strategic management research. Our replication study demonstrates the consequences of this inattention – including the underreporting of effects and potential for Type II errors.

The focus of most concern has been to avoid Type I errors, yet Type II errors also deserve attention. Important relationships may exist and go undetected. These relationships may be particularly important in the test of theory and interpretation of results for practice. In fact, Type II errors are just as critical as Type II errors in the effect on the development of our field and practice. And, Type II errors may be underreported in the published research because it is more difficult to publish research that report few statistically significant results. Thus, these results are particularly important for the strategic management field.

Our purpose is not to criticize prior work but to identify and emphasize needs for future research designs and methodologies in strategic management. While the field has developed significantly since Dan Schendel founded the *Strategic Management Journal*, our results emphasize the need for better empirical research to move the field forward; we may have reached another plateau in the development of the field. For the strategic management field to develop further and to mature into a well-respected field accepted by its sister social science disciplines, significant attention should be placed on measurement in strategic management research. Lest we seem overly critical of strategic management research, we would note that similar problems are present in other fields as well. A meta-analytic review of 70 studies from various social sciences concluded: "Measurement error, on average, accounts for most of the variance in a measure. This observation raises questions about the practice of applying statistical techniques based on the assumption that trait variance is large in relation to measurement error variance (Cote & Buckley, 1987: 317)." We strongly suspect that these concerns can be generalized to much research in the business disciplines. However, for the strategic management field to advance, it should not imitate its sister disciplines; rather it should take a leadership role in the conduct of high quality research.

We recommend that significant attention be paid to measurement in future strategic management research. To reduce measurement error, strategic management researchers should increase their concern for the construct validity of their measures. Measurement error can often be reduced by using multiple rather than single indicators for specific constructs as shown by the research presented herein. It is critical that steps be taken to insure high reliability of the measures used. Thus, we strongly encourage the application of more indexes and scales in strategic management research. Reliability tests should be considered in the research design and process. Doing so will likely require a break from past and current practices in the field. As we noted, a large number of studies in strategic management rely on measures used in published research. On the one hand, such practice enhances the ability to compare the findings with previous ones. However, given the serious measurement problems identified herein, relying on past research for measures ensures that the problems persist. Therefore, to use multiple indicators for constructs under study may require strategic management scholars to develop new measures with reliability as a major criterion.

Strategic management scholars should also display more sensitivity to the statistical power of their samples. The reduction of measurement problems may help to avoid debates as the one between Amihud and Lev and Lane et al. It may also help resolve conflicting findings thereby contribute to the resolution of important research questions. As such, it will enhance the power of strategic management research to contribute more value to the conversation involved in the practice of strategy by firms and top executives. While much of the practitioner literature, even in the better journals, such as *Harvard Business Review* and the *Sloan Management Review*, does not rely heavily on the scholarly research, there is a closer congruence between the research in strategic management and the practitioner literature on common topics. Therefore, if research on governance, for example, is to be used to formulate improved governance practices in organizations, indeed even base laws and regulations on it, we must be assured that the results of the research are accurate. If firms are to invest heavily to develop core competencies, the research should show that having stronger core resources leads to the achievement of a competitive advantage and importantly that we can have confidence that the relationship discovered by the research is correct. Thus, such changes will enrich the field and its value added contribution to knowledge of managing business enterprises.

In point of fact, the field is unlikely to make substantial progress without such attention. For example, extend the problems of interpreting results of the research on the relationship between governance controls and product diversification to other primary research areas in strategic management. Given that diversification is one of the most researched topics in the field, other areas are likely to be less well developed. Certainly it is more difficult to do high quality research on some content topics than on others. Some content areas may be better developed allowing more fine-grained measures to be built. Additionally, it may be easier to access data on some topics than others (allowing the selection of larger and higher quality samples and the identification of multiple measures). Yet, our research suggests that the measurement problems are relatively widespread in strategic management research. Thus, we can conclude that the measurement problems may be slightly easier to correct in some content areas than others.

Interestingly, we found indications that measurement problems were fairly consistent across journals regardless of their quality ranking and scholars regardless of the quality of their training (see Appendix A). Citations were largely unrelated to the quality of the measurement used (see Appendix A.1). In second tier journals, citations were partly related to the use of better measures. While there can be several interpretations for this outcome, perhaps, the quality of the journal is used as a general proxy for the assumed quality of the research. In second tier journals, the articles using higher quality measurement are more likely to be cited whereas, scholars in the field attribute high quality research to articles published in top tier journals (there is a reason they are ranked in the 'A' category), and question less the measurement issues.

Doing quality research is difficult and often requires significant amounts of effort and other resources and frequently takes considerable time. Thus, the 'publish or perish' mentality that pervades many academic institutions likely has a negative effect on the conduct of quality research. Certainly, their emphasis is largely on doing quality research but this is measured by publishing the work in the highest quality journals (based on their rankings primarily using their acceptance and citation rates). Earlier, we noted that quality of journal does not appear to be a good proxy for the quality of the measurement used in the research. Additionally, much of the research is conducted by scholars who are younger in the field, especially by untenured assistant professors. They must work under time pressures because of the limited time for tenure decisions in most universities and they have the fewest resources on which to call for accessing samples and data. Perhaps universities should focus a higher percentage of their resources on the tenure-track untenured scholars and lengthen the time limitations for tenure decisions. However, these actions are unlikely to lessen the problem substantially as further analysis revealed few differences in the quality of measurement employed by junior and senior faculty (see Appendix A.2). To reduce measurement problems and promote more effective research designs, reviewers and editors must adopt consistent and high standards in these areas. It is especially important for the gatekeepers with the top journals to take such actions because unless they do so, the quality of strategic management research is unlikely to change. Rewards (e.g., tenure, promotions, pay increases, endowed positions, teaching loads) are based on publishing in these journals; thus researchers will do what is necessary to publish in them. Promotion in the field is based on publication and the influence of those works. Supplemental analysis (see Appendix A.3) suggests that an articles' quality of measurement is largely unrelated to subsequent citation. The primary penalty for poor measurement may be the bottom drawer solution. Studies with poor measurement that do not yield significant results are not published, whereas studies with poor measurement that escape the gauntlet of measurement-induced type I error pay no penalty – they are published and subsequently cited.

Alternately, we are mindful of the "normal science straitjacket" to which Daft and Lewin (1990) referred. We recognize that some research in new areas may require a more flexible approach and standards in order to promote the research in this area (e.g., empirical research on emerging markets where primary and secondary data is difficult to obtain) because of its importance. Yet, we see these times as exceptions rather than common rules. Additionally, Daft and Lewin's (1990) intent was to promote more qualitative and alternative modes of research with the purpose of building theory and to provide richer data on which to more effectively interpret the results of large sample studies. This need continues to be of importance and thus we reemphasize it. Our research in no way diminishes this need. In fact, high quality qualitative research complements high quality quantitative research. In tandem, quality research of both types can move the field forward more rapidly.

In support of the conclusions noted above, Bergh (2001) suggested that future strategic management research is likely to place greater emphasis on research designs, construct validation and newer and more sophisticated analytical strategies. While building strong theoretical bases is highly important, measurement is at least equally important for future advances in the field of strategic management. We hope that this work serves as a catalyst to this end.

APPENDIX A

We conducted a series of post hoc, exploratory supplemental analyses to provide insight into possible causes and consequences of measurement error.[8] We examined two aspects potentially related to the use of quality measurement: journal effects and author characteristics. We also examined the effect that measurement quality has on subsequent citation.

Exploring these factors required quantifying the quality of measurement at the article level. We calculated this via an ordinal approach. We summed the use of each measurement type (i.e., single indicators, scales, etc.) used for IV's, DV's, and controls for the article. We then calculated three overall article ratings based on the measures utilized.

First, we calculated an *ordinal weighted average rating* by assigning the measures with the greatest potential for assessing reliability higher values. We assigned the use of single items a value of 1, single ratios a 2, discrete items a 3, indexes a 4, and scales a 5. For example, if an article utilized 6 independent variables including 2 single items, 1 ratio, 1 index, and 1 scale, the article IV rating score was 2.16 [(2 X 1) + (1 X 2) + (1 X 4) + (1 X 5)] / 6). Higher values indicated more sophisticated measurement approaches with a range of 1 to 5 with a continuous distribution. We also rated articles based on the *most sophisticated* measurement utilized. This resulted in a value of 1 to 5 based on the single most sophisticated measure used. In the example above, the article would be coded as 5 because it included at least one scale. A third rating, a binary assessment of whether the article used *any multiple* measurement approaches (i.e., indexes or scales) was also calculated. Articles that used indexes or scales were coded as 1, those that did not were coded as 0.

A.1 Journal Effects

[8] We'd like to thank one of our reviewers for recommending an exploration of these issues.

We assessed potential journal effects by developing three ANOVA models to test for differences between the measurement approaches between the four journals in the sample (i.e., *SMJ*, *AMJ*, *ASQ* & *MS*). We utilized the *ordinal weighted average rating* as this is the most continuous of the measures with a range of 1.00 to 5.00. Results suggest no differences between the measurement approaches used between the journals for IV's (F = .442; sig. = .723), DV's (F = .965; sig. = .411), or control variables (F = 1.244; sig. = .296). None of the *post hoc* comparisons between specific journals were statistically significant. We cannot, however, rule out possible effects between a broader pool of outlets. The journals included in our sample were restricted to the highest echelon of publications outlets in the field (MacMillan, 1991). Differences between the quality of measurement within these journals may be less than compared to less prestigious outlets. However, these results suggest that outcomes from our overall study are not because of one or more outlier journals. Rather, the measurement approaches appear to be endemic to the field.

A.2 Author Characteristics

We examined three author characteristics: researcher skill, rank/seniority, and affiliation. We quantified the quality of author affiliation at the time of publication and the quality of author's degree granting institution (DGI) using Gorman ratings (a similar approach was found to be valid by Hitt, Bierman, Shimizu & Kochhar, 2001).

Two variants of each of these measures were used: the average for all authors and the rating of the "best" (e.g., most prestigious) among the authors. Higher quality institutions were coded with higher values. The rank / seniority of the authors was roughly proxied by the years elapsed since degree date for each of the article's authors. Both an average for all authors and the greatest time elapsed since award of degree were used in the analysis.

While caution is advised in interpreting these *post hoc* analyses, initial results suggest a limited relationship among these variables. Correlations were low and not statistically significant with one exception. The use of more sophisticated measurement approaches for dependent variables was significantly correlated with the rank / seniority of authors. It is important to note that this finding is not causal – the use of higher quality measurement approaches may alternatively be the cause or effect of rank / seniority.

A.3 Effect of Quality Measurement

We examined outcomes of quality measurement by examining its relationship with citation counts. We collected citation data for all of the articles using the SSCI database. Using the three ratings of article measurement quality, we examined correlations with the cumulative count of all citations to each article as listed in SSCI. This analysis was completed independently for IV's, DV's, and control variables.

We counted cumulative citations to articles by year for five years post publication. For articles published in 1998, a citation window from 1998 through most of 2003 was available (publication year to plus 5 years). For articles published in 1999 and 2000, the window was restricted to publication year plus 4 years and publication year to plus 3 years respectively.

We assessed the relationship between the quality of measurement within an article with subsequent citation counts using a liberal approach: correlations. As this approach excludes the effect of any covariates that may dampen the relationship, we can expect any association between the variables to be apparent.

The effect of poor measurement appears largely unrelated to the value of a study within the discipline. The relationship between the quality of measurement within an article and subsequent citations to the article is not statistically significant. Correlations ranged from .000 to

.176 and several were negative. Only two relationships were statistically significant at the $p < .05$ level: the relationship between the *most sophisticated* measurement scheme used for control variables for years 0 to 1 and for years 0 to 3. While this relationship is statistically significant, it is likely spurious. To accept this finding, we must conclude that authors cite works based on the quality of control variables but not based on the quality of the independent and dependent variables. Additionally, the use of multiple measurement approaches (i.e., indexes and scales) within control variables was extremely limited: only 2.6 percent of the control variables used these approaches.

While the overall relationship between measurement quality and citation appears to be nonsignificant, the relationship may vary based on the quality of the outlet within which the work is cited. To assess this, we separately counted citations in two of MacMillan's (1991) article pools. The "outstanding" pool includes *SMJ*, *AMR*, *AMJ*, *ASQ*, *MS*, and *Harvard Business Review*. The "acceptable" strategy journal pool includes *California Management Review*, *Interfaces*, *Journal of Business Strategy*, *Journal of Management*, *Journal of Management Studies*, *Long Range Planning*, *Organizational Dynamics*, and *Sloan Management Review*. We also coded citations identified in SSCI that were in any source other than the above as "other journals".

Results for independent variables indicate no relationship between the quality of measurement used to assess independent variables and citation in leading journals. However, there is a statistically significant relationship between the *most sophisticated* measure of independent variables used within a study and subsequent citation to the article in MacMillan's (1991) "acceptable" pool of strategy journals. No statistically significant relationships were found between measurement quality of independent variables and citation in other journals.

Similar results are found for the relationships between the quality of measurement of dependent and control variables. None of the relationships are statistically significant for citations in the pool of leading or other journals, whereas they are statistically significant for the citation in other **strategy** journals.

Interpreting these two relationships is somewhat problematic because, if both are accurate, we must accept that the quality of the measurement of variables is of less importance for studies that appear in what are presumably the disciplines' highest quality outlets than it is for those appearing in outlets of lesser quality. In a further post hoc analysis of the most cited articles, we examined a subset of articles with high citation counts in the "acceptable" pool to identify the use of the citations in the work. In the vast majority of cases, the citations were for the *results* of the prior work or, when cited for a methodological purpose, they were *not* associated with the more sophisticated measure (i.e., index or scale) appearing in the original article. We conclude from these analyses that citation of an article is largely unrelated to the quality of the measurement used.

REFERENCES

Acar, W., & Sankaran, K. (1999). The myth of unique decomposability: Specializing the Herfindahl and entropy measures? *Strategic Management Journal*, 20: 969-975.

Amihud, Y., & Lev, B. (1981). Risk reduction as a managerial motive for conglomerate mergers. *Bell Journal of Economics*, 12: 605-617.

Amihud, Y., & Lev, B. (1999). Does corporate ownership structure affect its strategy toward diversification? *Strategic Management Journal*, 20: 1063-1069.

Babbie, E. (1989). *The Practice of Social Research*, 5th ed. Wadsworth Publishing Company, Belmont, California.

Berger, P. G. & Ofek, E. (1995). Diversification's effect on firm value. *Journal of Financial Economics*, 37: 39-65.

Bergh, D. D. (2001). Diversification strategy research at a crossroads. In M.A. Hitt, R. E. Freeman and J.S. Harrison (Eds.), *Handbook of Strategic Management.* Oxford, UK: Blackwell Publishers.

Bergh, D. D. & Holbein, G. F. (1997). Assessment and redirection of longitudinal analysis: Demonstration with a study of the diversification and divestment relationship. *Strategic Management Journal*, 18: 557-571.

Berle, A. A., & Means, G. C. (1932). The Modern Corporation and Private Property. New York, N.Y.: Macmillan Co.

Blalock, H. M. (1979). Social Statistics (2nd ed.). New York: McGraw-Hill.

Bowen, H. B. & Wiersma, M. F. (1999). Matching method to paradigm in strategy research: Limitations of cross-sectional analysis and some methodological alternatives. *Strategic Management Journal*, 20: 625-636.

Boyd, B. K. (1994). Board control and CEO compensation. *Strategic Management Journal*, 15: 335-344.

Chatterjee, S. & Blocher, J.D. (1992). Measurement of firm diversification: Is it robust? *Academy of Management Journal*, 35: 874-888.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates. Second edition.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112: 155-159.

Cote, J.A., & Buckley, M.R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, 24: 315-318.

Daft, R.L. & Lewin, A.Y. (1990). Can organization studies begin to break out of the normal science straightjacket? An editorial essay. *Organization Science*, 1:1-9.

Denis, D. J., Denis, D. K. & Sarin, A. (1997). Agency problems, equity ownership, and corporate diversification. *Journal of Finance,* 52: 135-160.

Denis, D. J., Denis, D. K. & Sarin, A. (1999). Agency theory and the influence of equity ownership structure on corporate diversification strategies. *Strategic Management Journal*, 20: 1071-1076.

Eisenhardt, K., 1989. Agency theory: An assessment and review. *Academy of Management Review*, 14: 57-74.

Fama, E. F. & Jensen, M. C. 1983. Separation of ownership and control. *Journal of Law and Economics*, 26, 327-349.

Ferguson, T. D. & Ketchen, D. J. (1999). Organizational configurations and performance: The role of statistical power in extant research. *Strategic Management Journal*, 20: 385-395.

Franke, R. H., Edlund, T. W. & Oster, F., III. (1990). The development of strategic management: Journal quality and article impact. *Strategic Management Journal, 11*(3), 243-253.

Godfrey, P.C., & Hill, C. W. L. (1995). The problem of unobservables in strategic management research. *Strategic Management Journal*, 16: 513-533.

Hitt, M. A., Gimeno, J., & Hoskisson, R. E. (1998). Current and future research methods in strategic management. *Organizational Research Methods*, 1: 6-44.

Hitt, M.A., Bierman, L., Shimizu, K. & Kochhar, R. 2001. Direct and moderating effects of human capital on strategy and performance in professional service firms: A resource-based perspective. *Academy of management Journal,* 44: 13-28.

Hoskisson, R. E., Hitt, M. A., Johnson, R. A., & Moesel, D. D. (1993). Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management Journal*, 14: 215-235.

Hubbard, R., Vetter, D. E., & Little, E. L. (1998). Replication in strategic management: Scientific testing for validity, generalizability and usefulness. *Strategic Management Journal*, 19: 243-254.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structures. Journal of Financial Economics, 3(4), 305-359.

Jowett, B. (1999). *Plato: The republic*. Barnes & Noble Books.

Ketchen Jr., D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. Strategic Management Journal, 17(6), 441-460.

Kuhn, T. S. (1996). *The structure of scientific revolutions.* Chicago: University of Chicago Press. Third edition.

Lane, P. J., Cannella, A. A., & Lubatkin, M. H. (1998). Agency problems as antecedents to unrelated diversification: Amihud and Lev reconsidered. *Strategic Management Journal*, 19: 555-578.

Lane, P. J., Cannella, A. A., & Lubatkin, M. H. (1999). Ownership structure and corporate strategy: One question viewed from two different worlds. *Strategic Management Journal*, 20: 1077-1086.

MacCallum, R.C., & Browne, M.W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114: 533-541.

MacMillan, I. C. (1989). Delineating a forum for business policy scholars. *Strategic Management Journal, 10*, 391-395.

MacMillan, I. C. (1991). The emerging forum for business policy scholars. *Strategic Management Journal, 12*(2), 161-165.

Maxwell, S.E. (1980). Dependent variable reliability and determination of sample size. *Applied Psychological Measurement*, 4: 253-260.

Mazen, A., Magid, M., Hemmasi, M., & Lewis, M. (1987). Assessment of statisticalpower in comtemporary strategy research. *Strategic Management Journal*, 8: 403-410.

McEachern, W. A. (1975). *Managerial control and performance*. Lexington: Lexington Books.

Mone, M. A., Mueller, G. C., & Mauland, W. (1996). The perceptions and usage of statistical power in applied psychology and management research. *Personnel Psychology*, 49: 103-120.

Palepu, K. (1985). Diversification strategy, profit performance, and the entropy measure. *Strategic Management Journal*, 6: 239-255.

Palmer, J. P. (1973). The profit performance effect on the separation of ownership from control in large U.S. industrial corporations. *Bell Journal of Economics*, 4: 293-303.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, 61: 473-485.

Tahai, A. & Meyer, M. J. (1999). A revealed preference study of management journals' direct influences. *Strategic Management Journal, 20*(3), 279-296.

Venkatraman, N. & Grant, J. H. (1986). Construct measurement in organizational strategy research: A critique and proposal. *Academy of Management Review*, 11: 71-87.

Zimmerman, D. W. & Williams, R. H. (1986). Note on the reliability of experimental measures and the power of significance tests. *Psychological Bulletin*, 100: 123-124.

TABLE 1

METHODOLOGICAL CRITIQUES OF STRATEGIC MANAGEMENT RESEARCH

Study	Focus	Sample Size	Journals Reviewed	Time Frame	Journal Pool*	General Findings
Short, Ketchen & Palmer (2002)	Sampling	437 studies	5	1980-1999	All	Less than 20% of studies used random sampling; only 40% of studies checked for sample representativeness.
Bergh & Fairbank (2002)	Measurement of change	126 studies	1	1985-1999	All	Strategy researchers tend not to recognize methodological requirements while measuring changes; the typical approach used is usually inappropriate and could lead to inaccurate findings and flawed conclusions.
Bowen & Wiersema (1999)	Cross-sectional designs	90 studies	1	1993-96	Not reported	Insufficient attention given to common issues associated with cross-sectional designs
Ferguson & Ketchen (1999)	Power in configuration research	24 studies	6	1977-96	Not reported	92% of published papers in this research stream had insufficient power
Hubbard, Vetter & Little (1998)	Replications	37 studies	9	1976-95	Subset	Few replication studies published; replications more common in *SMJ* than *AMJ* or *ASQ*
Bergh & Holbein (1997)	Longitudinal designs	203 studies	1	1980-93	All	More than 90% of studies had Type I bias due to insufficient attention to methodological assumptions.
Ketchen & Shook (1996)	Cluster analysis	45 studies	5	1977-93	All	Implementation of cluster analysis methodology often less than ideal
Mone, Mueller & Mauland (1996)	Statistical power	210 studies	7	1992-94	Subset	Average statistical power of management studies is low, especially for small and medium effect sizes

* 'Subset' indicates that a sample of relevant articles were used; 'All' indicates that all papers meeting the study criteria/focus were included

TABLE 2

DESCRIPTIVE STATISTICS

	du	***dr***	**Sales**	**Assets**	**Equity**	**Duality**	**Dir. Pay**	**Dir. Equity**	**Owner Reps**	**Insiders**
1. *du*	1.00									
2. *dr*	0.12	1.00								
3. Sales	0.12	0.16	1.00							
4. Assets	0.01	0.06	0.69	1.00						
5. Equity	0.07	0.17	0.80	0.80	1.00					
6. Duality	0.10	0.07	0.06	0.02	0.06	1.00				
7. Director Pay	0.16	0.17	0.45	0.38	0.44	0.06	1.00			
8. Director Equity	-0.09	-0.09	-0.20	-0.24	-0.25	-0.26	-0.21	1.00		
9. Owner Reps	-0.07	-0.05	-0.15	-0.20	-0.23	-0.19	-0.23	0.52	1.00	
10. Insiders	0.04	-0.03	-0.05	-0.20	-0.16	-0.11	-0.15	0.12	0.22	1.00
X	0.29	0.15	7.47	7.63	6.48	0.79	21847	4.47	0.98	0.28
σ	0.41	0.28	1.09	1.44	1.23	0.42	9163	11.52	1.60	0.14

Note: Correlations greater than 0.08 significant at p=.05; values greater than 0.10 at p=.01.

TABLE 3

COMPARISON OF ALTERNATE MODELS

Model	Firm Size		Board Control		R^2
	γ	*t*-value	γ	*t*-value	
1. Full model (Figure 1)	.11	2.3	-.16	2.2	.25
2. Single indicator for diversification, multiple indicator for control					
DU	.09	2.0	-.15	2.1	.024
DR	.14	3.1	-.12	1.7	.034
Mean for set 2:			-*.14*		*.029*
3. Single indicator for diversification, single indicator for control					
(a) **DU** and…					
DUAL	.11	2.8	.09	2.2	.022
DIRCOMP	.06	1.3	.13	2.9	.027
PCTDIR	.11	2.5	-.07	1.6	.018
OWNGRP	.11	2.6	-.05	1.3	.017
INSIDER	.12	3.0	.05	1.2	.016
Mean for set 3(a):			*.08*		*.020*
(b) **DR** and…					
DUAL	.16	3.9	.06	1.5	.030
DIRCOMP	.11	2.4	.12	2.7	.038
PCTDIR	.15	3.7	-.06	1.3	.030
OWNGRP	.16	3.9	-.03	0.7	.027
INSIDER	.16	4.0	-.02	0.5	.027
Mean for set 3(b):			*.06*		*.030*
Mean for set 3 overall:			*.07*		*.025*

Note: Highlighted text denotes nonsignificant path estimates.

TABLE 4

NORMALITY OF SAMPLE SIZE AND VARIABLE DISTRIBUTIONS

		Distribution for					
		Sample Size (N)	**Independent Variables**	**Control Variables**	**Ratio of *N* to Independent Variables**	**Ratio of *N* to Control Variables**	**Ratio of *N* to (Independent + Control Variables)**
Mean		2,559.22	5.66	8.62	628.70	606.69	236.12
s.d.		12,908.56	5.09	30.26	3,097.24	3,537.90	1,333.82
Minimum		20.00	1.00	0.00	2.80	1.00	1.00
Maximum		158,782.00	36.00	507.00	39,695.50	39,695.50	19,847.80
Skewness		8.42***	2.29*	10.47***	8.78***	8.91***	11.73***
Kurtosis		83.41***	7.44***	139.33***	94.00***	86.08***	161.41***
Percentiles:	25^{th}	98	2	0	23	22	12
	50^{th}	215	4	3	58	39	24
	75^{th}	426	7	8	142	84	65
	85^{th}	825	10	10	236	208	112
	90^{th}	1,461	12	15	362	275	179
	95^{th}	8699	15	18	1,906	1,477	597

Note: Significance of skewness and kurtosis based on Z scores with: * $p < .05$; ** $p < .01$; *** $p < .001$

TABLE 5

RESULTS OF CONTENT ANALYSIS

	IV's Count	IV's % of Variables	IV's % of Tests	DV's Count	DV's % of Variables	DV's % of Tests	Controls Count	Controls % of Variables	Controls % of Tests
Total Number of Variables	3,388			677			5,376		
Total Number of Tests	625			625			625		
Single Indicators									
# of single indicators used	1,613	47.6%		258	38.1%		4,280	79.6%	
# of tests using single indicators	335		53.6%	238		38.1%	371		59.4%
of the tests, # that rely solely on single Indicators as IV, DV, <u>or</u> control	127		20.3%	170		27.2%	176		28.2%
of the tests, # that rely exclusively on single indicator for IV, DV <u>and</u> control	34		5.4%	34		5.4%	34		5.4%
Single Ratios									
# of single ratios used	309	9.1%		101	14.9%		435	8.1%	
# of tests using single ratio indicators	119		19.0%	95		15.2%	146		23.4%
of tests, # that rely solely on single ratios as IV, DV, <u>or</u> Control	36		5.8%	64		10.2%	27		4.3%
of the tests, # that rely exclusively on single ratios for IV, DV <u>and</u> control	0		0.0%	0		0.0%	0		0.0%
Discrete Items									
# of discrete items used	479	14.1%		32	4.7%		268	5.0%	
# of tests using discrete indicators	96		15.4%	32		5.1%	36		5.8%
of the tests, # that rely solely on single ratios as IV, DV, <u>or</u> control	36		5.8%	31		5.0%	15		2.4%
of the tests, # that rely exclusively on single ratios for IV, DV <u>and</u> control	0		0.0%	0		0.0%	0		0.0%
Indexes									
# of indexes used	502	14.8%		117	17.3%		103	1.9%	
Average reliability	.86			.74			.74		
# of tests using indexes	162		25.9%	114		18.2%	64		10.2%
of tests using indexes, # reporting reliabilities	33		20.4%	4		3.5%	9		14.1%
of the tests, # that rely solely on index as IV, DV, <u>or</u> control	47		7.5%	78		12.5%	0		0.0%
of the tests, # that rely exclusively on index for IV, DV <u>and</u> control	0		0.0%	0		0.0%	0		0.0%
Full Scales or Multiple Indicators									
# of scales used	399	11.8%		133	19.6%		39	0.7%	
Average reliability	.80			.82			.76		
Mean # of indicators	6			6			5		
# of tests using scales	171		27.4%	100		16.0%	26		4.2%
of tests using scales, # reporting reliabilities	133		33.3%	84		63.2%	16		41.0%
of the tests, # that rely solely on scale as IV, DV, <u>or</u> control	62		9.9%	79		12.6%	3		0.5%
of the tests, # that rely exclusively on scale for IV, DV <u>and</u> control	2		0.3%	2		0.3%	2		0.3%
None used							184	3.4%	
Not specified/detail missing	86	2.5%		36	5.3%		67	1.2%	
		100.0%			100.0%			100.0%	

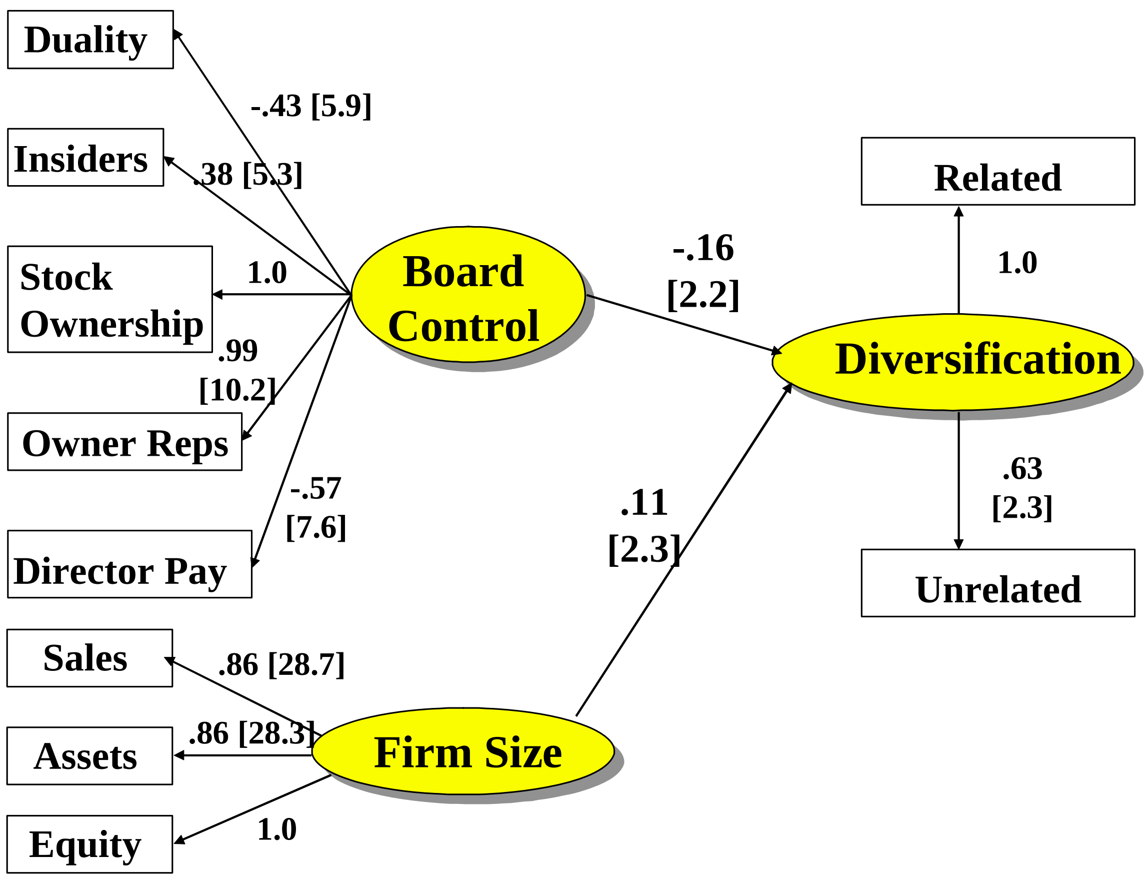
TABLE 6

COMPARISON OF THE USE OF SINGLE AND MULTIPLE MEASUREMENT APPROACHES

		IV's			*DV's*			*Controls*	
	Count	**% of Variables**	**% of Tests**	**Count**	**% of Variables**	**% of Tests**	**Count**	**% of Variables**	**% of Tests**
Total Number of Variables	3,388			677			5,376		
Total Number of Tests	625			625			625		
None used							184	3.4%	
Not specified/detail missing	86	2.5%		36	5.3%		67	1.2%	
Measurement Approach Not Allowing for Assessment of Reliability (single, ratio & discrete measures)									
# of single, ratio, and discrete measures used	2,401	70.8%		391	57.7%		4,983	92.7%	
# of tests using single, ratio, and discrete measures	550		88.00%	365		58.40%	553		88.48%
of the tests, # that rely solely on single, ratio, or discrete for IV, DV, or control	199		31.84%	265		42.40%	218		34.88%
of the tests, # that rely exclusively on single, ratio, or discrete for IV, DV and control	34		5.44%	34		5.44%	34		5.44%
Measurement Approach Allowing for Assessment of Reliability (indexes and scales)									
# of indexes and scales used	901	26.6%		250	36.9%		142	2.6%	
Average reliability reported	0.83			0.78			0.75		
# of tests using indexes or scales	333			214			90		
# reporting reliabilities	166		49.85%	88		41.12%	25		27.78%
of the tests, # that rely solely on indexes or scales for IV, DV, or control	109		17.44%	157		25.12%	3		0.48%
of the tests, # that rely exclusively on indexes or scales for IV, DV and control	2		0.32%	2		0.32%	2		0.32%

FIGURE 1

RESULTS OF STRUCTURAL MODEL

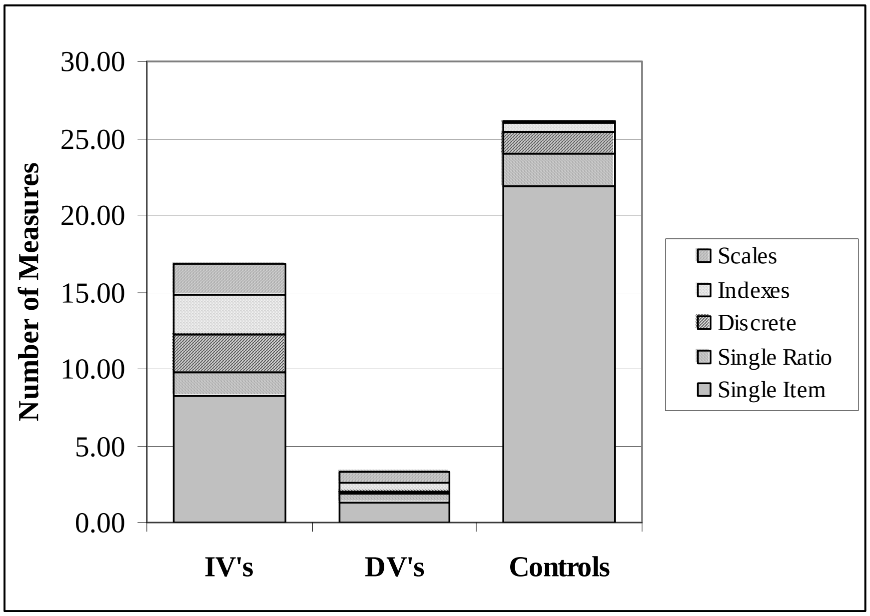


Note: Certain terms (e.g., theta and phi matrices) omitted for ease of representation. T-values of parameters noted in brackets; significance levels as follows: t = 2.0, p< .05; t = 2.7, p< .01; t= 3.5 p < .001.

FIGURE 2

ARTICLE LEVEL MEASUREMENT

A. Frequency of Measurements Used Within Typical Article



B. Ability to Assess Reliability Within Typical Article