The Fortune Corporate ‘Reputation’ Index: Reputation for What?

Gerald E. Fryxell
University of Tennessee

Jia Wang
California State University-Fresno

This study examines competing measurement models of the Fortune reputation data and argues that, for research purposes, its usefulness is limited to measuring the extent to which a firm is perceived as striving for financial goals. The generalizability of this conclusion is evaluated and past research studies purporting to have measured other constructs (e.g., social responsibility, quality, stakeholder orientation) are revisited.

Few social scientists would dispute that competent research is built on a foundation of good measurement. However, in many areas of management study, measurement is a particularly vexing problem. Fortunately, some progress has been made. For example, a reliance on single exemplars of complex constructs and the exclusive use of self-report data are increasingly being challenged (e.g., Hughes, Price, & Marrs 1986; Venkatraman & Grant, 1986). Nevertheless, in some areas of management research such as corporate social responsibility, innovation, and quality, measurement concerns cut to the heart of entire research streams (e.g., the relationship between social responsibility and profitability—see Aupperle, Carroll, & Hatfield, 1985 for a review).

From time to time, a data source emerges which portends to offer a solution to an embarrassing trail of equivocal findings. A recent example of this may be found in the Fortune “reputation” data, which has recently been used to measure such disparate and problematic constructs as: a firm’s stakeholder orientation (Chakravarthy, 1986; Preston & Sapienza, 1990); corporate social responsibility (Conine & Madden, 1986; McGuire, Sundgren, & Schneeweiss, 1988); management quality (McGuire, Schneeweiss, & Branch, 1990); and corporate reputation (Fombrun & Shanley, 1990). Ostensibly, this survey has several beguiling features; ratings from some 8,000 experts; a rank ordering approach within a competitor set allowing some control of industry effects; and a 50 percent response rate.
In spite of these admirable characteristics, sufficient information has aggregated over time which we believe casts considerable doubt on the usefulness of the Fortune data to capture many of these constructs. First, both Conine and Madden (1986) and Preston and Sapienza (1990) reported high correlations between measures of financial performance within the survey and the single Fortune item pertaining to a firm’s responsibility to community and environment. Given the equivocality of this relationship in the literature, the strength of these findings casts some suspicion that other influences (e.g., mono-method bias or “halo” effect) may have played a significant role in linking these constructs. Second, using the same single social responsibility item, McGuire et al. (1988) revealed some difficulty in mustering evidence of its validity by resorting to a rather modest correlation (.45) between a subset of 58 firms in their sample and independent ratings by the Council on Economic Priorities. Moreover, in a more recent study by McGuire et al. (1990) the pattern of correlations between the eight items in the Fortune survey and ten financial measures is strikingly similar. This might suggest that some larger general factor is influencing each item (which given the background of the raters could very well be grounded in the investment potential of each firm). Relatedly, Fombrun and Shanley recently reported two statistics pointing to the possibility that the survey may be unidimensional—an unusually high reliability statistic of .97 and the extraction of a single factor which “accounted for 84 percent of the variance” (1990, p. 245). Most recently, Fortune attempted a rather ineffective defense of the virtue of their survey, proclaiming that “roughly half of the overall reputation index “can be explained by a company’s financial performance over time” (1993, p. 44). While they appear sanguine that any remaining variance is due to other “subjective factors,” the intrusion of measurement error in a typical survey could sufficiently erode this residual so as to leave little room for other constructs.

Unidimensionality would suggest that items from this survey cannot be used singly or in combination to measure such disparate constructs as innovativeness, social responsibility, wise use of corporate assets, or the quality of a firm’s products. Thus, it seems likely that in spite of its popularity and internal consistency, the Fortune reputation survey may have been misused. The purpose of this study is to investigate this possibility by examining the measurement structure of the Fortune data through an evaluation of several competing measurement models.

Methodology

The Fortune Ratings

Fortune data used in this study were collected for the years 1985 to 1989 inclusive, although data from the year 1986 is used for the bulk of this analysis since, at the time the data were obtained, it was in the middle range of years present in other studies. In Fortune’s survey, each company is rated relative to its leading competitors on eight characteristics using an 11-point scale (0 = poor, 10 = excellent) on the following eight attributes (with abbreviations):
Long-term investment value (L-TIV); Financial soundness (FINS); Wise use of corporate assets (WUCA); Quality of management (QOM); Quality of products or services (QOP/S); Innovativeness (INN); Ability to attract, develop, and keep talented people (ATP); and, Community and environmental responsibility (CER). A somewhat more detailed discussion of this survey can be found in McGuire et al. (1988).

**Analysis of Competing Models**

The relationship among the eight *Fortune* items and its underlying construct(s) were examined by positing competing theoretical models and then evaluating their conformance with the data using confirmatory factors analysis within the LISREL program (Jöreskog & Sörbom, 1989). The three substantive models evaluated are shown in Figure 1. In addition, a null model, which is necessary to calculate certain measures of model fit, was also estimated. As it is more meaningful to compare the performance of firms within their industry grouping, we normalized all of the items relative to industry means. This transformation yielded a correlation matrix that was subsequently analyzed by the program.

**Null Model** ($M_0$). This is a model which posits that each item perfectly and orthogonally measures a unique factor. The objective is to obtain a baseline $\chi^2$ fit statistic reflecting a model where a priori knowledge about its factor structure would not be assumed (i.e., every item is a distinct factor—Bentler & Bonnet, 1980; Widaman, 1985). In matrix terms, this is a model where $\Lambda_0$ is a null matrix, $\Phi$ is an identity matrix, and $\Phi_0$ is a diagonal matrix.

**A Single-Factor Model** ($M_1$). The previously mentioned statistics reported by Fombrun and Shanley (1990), and indeed their implicit use of all eight *Fortune* items to create an index of corporate reputation, imply a model where all items load on a single factor. Moreover, the *Fortune* data include an average overall rating which encourages the implicit adoption of this model (e.g., Chakravarthy, 1986; McGuire et al., 1990). Consequently, this model has been both empirically and conceptually suggested. In this model each item loads on a single reputation factor, the factor's variance is set to unity, and the error variances are freely estimated.

**A Two-Factor Model** ($M_2$). Carroll conceptualized the responsibilities of a firm as "a range of obligations business has to society" (1979, p. 499). Although he argued for four ordered categories of obligations (i.e., economic, legal, ethical, and discretionary), these categories are not well represented by the items in this survey. Legal obligations are certainly not in evidence among the *Fortune* items and the CER item appears to be the lone exemplar from either the ethical or discretionary domain. Indeed, given the composition of the *Fortune* raters and the content of the item, an alternative view of the CER item could be that it represents a type of stakeholder management capability where community and environmental interests are well managed so as to facilitate strategy implementation (similar to the "stakeholder orientation" perspective employed by Chakravarthy, 1986 and Preston & Sapienza, 1990). Moreover, a key concern of many social issues researchers has been to achieve...
FRYXELL AND WANG

Model 1—Single Factor Model

(SUBJECT TO MAJOR REVISION)

Model 2—Two Factor Model

(Figure 1 continued)
Model 3—Dominant Factor Model

Legend:  
\(^a\) The symbol "\(e\)" is used instead of "\(\delta\)" to represent error and the use of other LISREL parameters has been omitted to facilitate interpretation.  
\(^b\) The arc indicates that the factors are intercorrelated.  
\(^c\) This model will be estimated both with and without the CER item permitted to load on Dominant Financial factor.

Figure 1. Competing Measurement Models of FORTUNE Items

a greater integration of social responsibility and economic performance (e.g., Wartick & Cochran, 1985; Wood, 1991). Thus, conceptually it is difficult to argue that all of the items do not pertain to economic responsibilities in some way (although based on the equivocality of empirical research, the link of the CER item to financial goals is admittedly more tenuous than for the other process items).

Based on the brief content wording of the items, the most salient qualitative difference from the perspective of the Fortune rater set (i.e., executives, directors, and financial analysts) is a means/ends distinction. As the first three items are so clearly financial in their content, from the rater’s point of view they appear to map onto basic beliefs in the primacy of attaining financial, growth and risk targets in economic organizations. More theoretically, this is also the economic subset of Wartick and Cochran’s (1985) view of social “principles” (i.e., social obligations); Preston and Sapienza (1990) also linked these three items together under the rubric of a firm’s “shareholder-orientation.” In this context, the remaining items appear to capture capabilities or means related to the accomplishment of these ends. Thus, a model is proposed where the strictly financial items (i.e., L-TIV, FINS, and WUCA items) load on a single “financial ends” factor, whereas all other items load on a second factor which we will call “capabilities and strategic means”. This appears to fit well within Clarkson’s operationalization of this component of the model as “how
an organization responds to changing values, issues and conditions...” (1988, p. 246). Due to the obvious connection between “capabilities” and “financial goals,” we would expect these two factors to be strongly correlated.

A Dominant-Factor Model (M3). This model modifies the previous two-factor model (M2) by recognizing that the capabilities and strategic means items are instrumentally linked to financial ends such that respondents may cognitively carry over evaluations of quantitative and published information on financial performance to the other items (i.e., a halo effect—considered by many to be the major psychometric error effecting multifactor rating, see Cooper, 1981). Consequently, this model proposes two factors (i.e., “dominant financial” and “subordinated means”) where the process items load on both factors. Thus, this modification recognizes both substantive and methodological possibilities. In order to further explore this issue and recognize the uniqueness of the CER item, two variants of this model were estimated depending on whether or not the CER item is allowed to load on the dominate factor (i.e., in M3a it does not, whereas in M3b this is permitted). This seems warranted as Aupperle et al. (1985) empirically demonstrated evidence of an inverse relationship between economic and ethical dimensions in an exploratory factor assessment of items using a different forced-choice survey instrument.

Subgroup Analysis

Assuming that one model is demonstrated to have a superior fit with the data, additional analyses will evaluate the stability and generalizability of the model. First, comparisons will be made using data from 1985 to 1989 on the same companies averaged over a five year period. The purpose of this comparison is to assess the model’s stability over time. The second comparison will be made between firms which were consistently evaluated over the five year period to those firms which were only occasionally rated during this time. This acknowledges the possibility that companies which are entrenched members of the Fortune “club” may reflect a somewhat different measurement structure. For example, raters may simply have more qualitative information available on the more entrenched firms. Fombrun and Shanley (1990), for example, found that visibility (measured by the number of articles on a company) was significantly related to their index of reputation.

Results

Table 1 presents the correlations among the eight Fortune items after their normalization to industry means. The confirmatory factor models were estimated using the LISREL program; the selected fit statistics and indices are reported in Table 2. The fit statistics given are those most commonly reported (e.g., Harvey, Billing, & Nilan, 1985; Marsh & Hocevar, 1985). However, in order to arrive at the most acceptable model, we adhered to a procedure using nested comparisons which relied on $\chi^2$ difference tests to evaluate the effect of incremental changes. An initial comparison of $M_1$ and $M_2$ showed a $\chi^2$ change of 46.45 when the two factors in $M_2$ were merged (by fixing their intercorrelation...
### Table 1. Correlation Matrix of FORTUNE Survey Items

(n = 292)

<table>
<thead>
<tr>
<th>FORTUNE ITEM</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Long-Term Investment Value (L-TIV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Financial Soundness (FIN)</td>
<td>.868</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Wise Use of Corporate Assets (WUCA)</td>
<td>.909</td>
<td>.851</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Quality of Management (QOM)</td>
<td>.929</td>
<td>.821</td>
<td>.924</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Quality of Products/Services (QOP/S)</td>
<td>.763</td>
<td>.729</td>
<td>.714</td>
<td>.796</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6) Innovativeness (INN)</td>
<td>.788</td>
<td>.636</td>
<td>.729</td>
<td>.796</td>
<td>.687</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7) Ability to Attract/Develop/Keep Talented People (ATP)</td>
<td>.899</td>
<td>.849</td>
<td>.852</td>
<td>.922</td>
<td>.843</td>
<td>.801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) Community and Environmental Responsibility (CER)</td>
<td>.635</td>
<td>.629</td>
<td>.545</td>
<td>.641</td>
<td>.709</td>
<td>.561</td>
<td>.735</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

* All variables were normalized with respect to industry group means. Consequently all have zero means and unitary standard deviations.

* All correlations are significant at $p < .001$.

### Table 2. $\chi^2$ Differences for Nested Model Comparisons and other Fit Indices for Models of FORTUNE Survey Items

<table>
<thead>
<tr>
<th>MODEL</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$\Delta \chi^2$</th>
<th>Rho</th>
<th>$\Delta$</th>
<th>PFI</th>
<th>RMSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) M₀—Null Model</td>
<td>3225.12</td>
<td>28</td>
<td>—</td>
<td>.880</td>
<td>.909</td>
<td>.670</td>
<td>.040</td>
</tr>
<tr>
<td>2) M₁—Single “Reputation” Factor Model</td>
<td>295.82</td>
<td>20</td>
<td>46.45/1df</td>
<td>.895</td>
<td>.923</td>
<td>.646</td>
<td>.035</td>
</tr>
<tr>
<td>3) M₂—Two Factor Model</td>
<td>249.37</td>
<td>19</td>
<td>104.99/4df</td>
<td>.924</td>
<td>.955</td>
<td>.528</td>
<td>.025</td>
</tr>
<tr>
<td>4) M₃ₐ—Dominant Factor Model w/out CER Item</td>
<td>144.38</td>
<td>15</td>
<td>10.10/1df</td>
<td>.918</td>
<td>.955</td>
<td>.493</td>
<td>.025</td>
</tr>
<tr>
<td>5) M₃₉—Dominant Factor Model w/CER Item</td>
<td>144.28</td>
<td>14</td>
<td>10.10/1df</td>
<td>.918</td>
<td>.955</td>
<td>.493</td>
<td>.025</td>
</tr>
</tbody>
</table>

**Note:**

* The $\chi^2$ difference test between nested models indicates whether a significant improvement fit is associated with the free estimation of the parameter changes. In this instance: model M₂ would be accepted over M₁ and M₃ₐ over M₂; however, M₃₉ would not be accepted over model M₃₉.
to unity with a gain of one degree of freedom). As this change revealed a much poorer fit of the single factor model ($M_1$), it was rejected in favor of the two factor model ($M_2$). Similarly, $M_2$ was compared to $M_{3a}$ and the incremental change to the less restrictive model was accompanied by a large $\chi^2$ drop of 104.994df. Thus $M_{3a}$ was easily accepted over both $M_1$ and $M_2$. Model $M_{3b}$, however, afforded virtually no improvement of fit over $M_{3a}$ with the loss of a degree of freedom. Consequently, $M_{3a}$ is clearly the "accepted" model on the basis of these comparisons. As is often the case, however, there is a disagreement among the other fit indices (Harris & Schaubroeck, 1990); however, in this instance the only index which point away from $M_{3a}$ is the parsimonious fit index (PFI), which favors models which preserve degrees of freedom among similar models and is somewhat less appropriate given the dissimilarity of these models (James, Mulaik, & Brett, 1982).

The parameter estimates for each of these models are reported in Table 3. For the ease of presentation, we have chosen to exclude the error variances from this table, although the values of all of these variances were quite reasonable (in particular, negative variance anomalies—a possibility with maximum likelihood estimates—were not encountered). While the parameters for $M_{3a}$, the accepted model, should be the focus of attention, we found it instructive to observe the relative changes in the loadings as we moved toward that model.

The single-factor model reveals uniformly high loadings form .969 ($\lambda_{4,1}$) to .686 ($\lambda_{8,1}$). The pattern of loadings reveals that two of the three highest loading items pertain to staffing issues (i.e., QOM and ATP—$\lambda_{4,1}$ and $\lambda_{7,1}$) which are intermingled with the three financial items. However, even the QOP/S, INN, and CER items ($\lambda_{5,1}$, $\lambda_{6,1}$, and $\lambda_{8,1}$) show strong linear relationships with this factor. Such results might suggest that the Fortune items seem to measure the same construct with a high degree of consistency and are further suggestive that its conceptual domain is fairly broad (this is essentially the conclusion reported by Fombrun & Shanley, 1990). However, our results show that such a model should be rejected in favor of a two factor model.

In moving to a two factor model ($M_{2}$), although the first three items retain their high loadings on the "financial ends" factor, the remaining items now appear to load strongly on a different factor. As the highest loadings on the "capabilities and strategic means" factor pertain to quality of management and other personnel, the raters would appear to be evaluating overall talent. While both sets of loadings would appear to be candidates for a reasonably good index, this potential, however, is clearly betrayed by the extremely high intercorrelation (.971) between the two factors.

The accepted dominant-factor model ($M_{3a}$) substantially attenuates this intercorrelation and reveals changes in the composition of the first factor representing the raters’ perception that the firm meets its financial goals. In this model, while the financial items clearly maintain prominence in their linear relationship to the dominant factor, the QOM item also loads strongly on this factor. Assuming a large number of the raters probably adhere to the common view that management has a primary fiduciary responsibility to the firm’s
owners, the first factor appears to be delimited to what Venkatraman and Ramanujam called “the domain of financial performance” (1986, p. 803). However, it is noteworthy that all of the process items with the exception of the CER item (QOP/S, INN, ATP—λ₅,₁, λ₆,₁, and λ₇,₁) have significant loadings on this factor as well. This, we think, is strongly suggestive of a halo effect where evaluations of the financial prowess of a company are projected onto these other items.

The second, subordinate factor has only two items (i.e., CER and QOP/S—λ₈,₂ and λ₉,₂) with loadings in excess of .5. We find it interesting that the *Fortune* raters should relate community and environmental responsibility to product/service quality. Along with the loadings of the other items, this suggests that this factor may pertain more to a broader stakeholder capability of the firm than to economic capabilities, per se. However, this model also reveals that these same items would comprise a poor index for the purpose of measuring such a construct. This weakness is evidenced by the observation that every item, with the exception of the CER item, is also directly, and usually heavily, influenced by the raters’ financial evaluation of the firm. Moreover, as the two factors are so highly intercorrelated (φ₂,₁ = .786 in M₃b), the subordinate factor appears to be strongly embedded in the first such that clear discrimination between the two constructs is unlikely. This is especially true since it is likely
that the raters had relatively vague conceptual clarity on what was probably a subsidiary concern.

Although estimates obtained for model $M_{3b}$ are problematic (i.e., possible under-identification of $\phi_{2,1}$), we found it interesting to observe that the CER item was negatively related to the first factor, which is similar to the finding of Aupperle et al. (1985). Thus, assuming there are halo effects here, they do not appear to extend to the CER item. While its loading on the second factor here exceeds unity, the same relative relationship to the other items is preserved. While the CER item’s “independence” may appear to raise the possibility that the CER item could be used by itself to measure “social responsibility” in some way, such an inclination should be resisted. Of course, should any financial constructs (or any of the other constructs suggested in the item content of the Fortune data) be also present, any estimates would clearly be biased. While the high level of intercorrelation exhibited by these two factors could be due to a variety of explanations, it must certainly be due to influences well beyond any actual level of association between these constructs which would then be unspecified in any model tested. It is our opinion that the most plausible explanation for this high level of intercorrelation is the presence of a portent mono-method bias in this database.

### Multiple Group Comparisons

We extended this evaluation to multiple group comparisons in order to assess the accepted model’s stability and generalizability. The first comparison involved ratings in a single year versus a five-year average; the second comparison was for firms which remained in the ratings for the entire five year period (but using the ratings only from 1986) versus those firms which were in and out of the ratings during the same period (but present in 1986). In both
comparisons two null hypotheses were tested as before using $X^2$ difference tests in nested models. The first test was of equivalence in factor loadings (i.e., invariance in the $A_x$ matrix involving 12 parameters); the second was a test of equivalence in the factor intercorrelations (i.e., the single parameter in the $\Phi$ matrix). The results of these tests are reported in Table 4.

In these tests we were unable to confirm any differences in the measurement structure either by time or by type of company. Consequently, these results show that the dimensionality of these ratings is reasonably stable; that is, the relationships demonstrated in this study should be generalizable to other time periods and to changes in the composition of the database.

Discussion

Taken together, these findings have important implications about what the Fortune data actually measure, or, stated differently, what these firms have a reputation for. We conclude that the dominant factor underlying the database appears to be predominantly financial in its construct domain. If such a measure should be sought by a researcher, then we would assert that the Fortune data provides a good measure (we would simply create an index of the first four items in our model). Indeed, such a measure would be an especially effective complement to other financial performance measures as part of a multiple measurement strategy. Beyond this rather limited use of the database, however, we have serious concerns. All but one item (i.e., the CER item) appear to be directly influenced by the raters' perceptions of the financial potential of the firm. Thus, researchers must be extremely wary so that the content of these items doesn't seductively lead them to believe they are measuring far more than they are.

This assertion about the narrowness of the construct's domain can be borne out simply by examining some of the previously reported correlations and regression relationships. Fombrun and Shanley (1990), for example, reported regression coefficients for an index of the eight Fortune items on thirteen objective variables gleaned from secondary sources. The significance patterns showed extremely strong relationships with profitability, risk, market-book ratio, and media visibility and weaker relationships with other variables. Also McGuire et al. (1990) reported a series of correlations between individual items and various financial indicators which produced strikingly similar patterns.

While this confirmatory factor analysis permits the technical isolation of a second factor, it is unlikely that a firm's reputation for specific capabilities or external-orientation can be sufficiently discriminated from its reputation for financial prowess to be usable. We believe this limitation also extends to the single CER item as the evidence presented here shows that it is also subject to a combination of mono-method bias and financial distraction. Stated differently, it seems highly unlikely that the Fortune's expert raters adequately discriminate between financial and nonfinancial aspects of a firm's reputation so as to permit their valid measurement. In light of these conclusions, it is instructive to reexamine some of the more salient previous uses of the Fortune data.
Reexamination of Previous Studies

Whereas Conine and Madden's (1986) conclusion (i.e., that corporate social responsibility and financial performance are positively related) was based solely on the Fortune data, it is clear that the common variance between the CER item and the financial ends factor demonstrated above (i.e., due to the high level of intercorrelation) would have, in itself, linked these two constructs. Consequently, the strength of their finding is more methodological artifact than confirmation of one of the more elusive relationships in social issues research (i.e., that social responsibility is also profitable).

McGuire et al. (1988) also reported findings relating the CER item to financial data; however, in this case they used secondary sources (e.g., COMPUSTAT). While this is undoubtedly an improvement over the previous study, their conclusion that prior firm performance is a better predictor of social responsibility than contemporaneous assessments of profitability must nevertheless be viewed with skepticism. In light of the likelihood of mono-method bias linking this item with appraisals of financial performance, a likely explanation for their findings would be that the stronger link to prior period performance is simply because rater are likely to base current perceptions of financial prospects on somewhat dated objective information. At a minimum, a reexamination of these relationships is warranted.

Preston and Sapienza (1990, p. 373) combined the features of both these studies by drawing conclusions about both relationships within the survey and relationships to secondary sources. Exemplary of the former, they concluded that, "stakeholder management'... appears to be fairly common among major U.S. companies" and, "major stakeholder groups apparently gain or lose from the same broad patterns of organizational success or difficulty." Based on correlations to secondary indicators of profit, growth and size, they also concluded that, "managers typically do not sacrifice major stakeholder objectives in pursuit of conventional growth goals" and so on. Clearly, the dominance of financial considerations which drive raters' evaluations evidenced in the accepted model appears to test the assumption that, "the data should be accepted at face value" (p. 368). Consequently, it appears likely that the ubiquitous nature of stakeholder managament that they are referring to has methodological origins in the ubiquitous nature of managing for profits.

More recently, McGuire et al. (1990) turned their attention to a "management quality" construct and, using a similar approach with the QOM item and the overall index, drew similar conclusions about the linkage between management quality and financial performance (e.g., that quality "reflects previous risk and return characteristics"; "the strength of the quality/ performance relation was stronger when evaluators focused on...innovation, financial soundness, use of assets," p. 178) Once again, the alternative explanation suggested by this study would be that these relationships may be spuriously linked to raters' perceptions of the companies' financial accomplishments.
Other researchers (e.g., Chakravarthy, 1986; Fombrun & Shanley, 1990) who have relied on the *Fortune* data as an overall and ostensibly reliable measure of a broader construct may also have erred, although to a much lesser degree, by imbuing it with excessive meaning. For example, Chakravarthy used the *Fortune* data as an indicator of firms that “appealed to multiple stakeholders” (1986, p. 448). Fombrun and Shanley, who initially raised doubts about the multidimensionality of the survey, appear to have ascribed an excessively broad content domain to their “reputation” construct and, as a result, may have overstated the dependence of a firm’s reputation on its “demonstrations for social concern” (1990, p. 252).

Regardless of how tightly this construct circumscribes financial performance, the most important point here is that researchers need to realize that this construct’s domain of meaning is much more delimited than previously assumed from simply inspecting item content. Part of the limitation of the *Fortune* survey to a financial construct is no doubt linked to its choice of “expert” raters and it is likely that quite different results would be obtained should the same instrument be used by a different group, although it is difficult to imagine what group’s opinions would be more highly valued.

In conclusion, this study has illuminated more clearly the measurement structure underlying the *Fortune* “reputation” data. In so doing, it appears that the “reputation” of these firms, as evaluated by this panel of industry expert speaks most directly to their “reputation” as an investment. Thus, we have raised serious doubts about the capability to measure the constructs implied by their item content due to an apparent interaction of mono-method bias and “halo” effects. In this light we have also gone well beyond the existing literature in demonstrating a degree of stability and generalizability of these relationships and reinterpreted findings from examples of previous work—some of which we think are seriously misleading. Beyond these specific contributions, however, we believe this study raises broader concerns about measurement and a need for some systematic evaluation of measures prior to their wide-spread implementation. This would appear to be especially true, as mentioned in the beginning of this study, in cases where good measures have been particularly troublesome. In this regard, we are reminded of the oft spoken cautions against the existence of the “free lunch.” Similarly, we believe that measures, such the *Fortune* data, that seem almost “too good to be true” would typically be those which warrant the biggest doses of skepticism.

**References**


