The Construct Validity of the Kinder, Lydenberg & Domini Social Performance Ratings Data

Mark Sharfman

ABSTRACT. Carroll (1991) encouraged researchers in Social Issues Management (SIM) to continue to measure Corporate Social Performance (CSP) from a variety of different perspectives utilizing a variety of different measures. In addition, Wolfe and Aupperle (1991) (and others) have asserted that there is no, single best way to measure CSP and that multiple measures and perspectives help develop the field. However, Pfeffer (1993) suggest that a lack of consistent measurement has constrained organization studies (and by implication, the field of social issues management,) in its development as a field. It may be in the best interest of social issues management researchers to try to develop a common body of measures and data. Recently, Kinder, Lydenberg and Domini & Co. (KLD – a social choice investment advisory firm) has made available their social performance database. The KLD data have potential to become a widely accepted set of CSP measures. The purpose of this paper is to present a construct validity study comparing the KLD data to other measures of CSP.

Carroll (1991) encouraged researchers in social Issues Management (SIM) to continue to measure Corporate Social Performance (CSP) from a variety of different perspectives utilizing a variety of different measures. Wolfe and Aupperle

Mark P. Sharfman is Assistant Professor of Strategic Management at the University of Oklahoma. His research interests are in corporate social performance, the firm/business environment relationship and measurement issues. He has published his research in the Academy of Management Review, Decision Sciences, Journal of Management and the Strategic Management Journal. He is a member the Academy of Management, International Association for Business and Society and Strategic Management Society.

survey and to data derived from the holdings lists of 11 “social choice” mutual stock funds.

Previous approaches to CSP measurement

Virtually all of the empirical CSP research (with CSP as either a dependent or an independent variable) can be classified into either single issue or more comprehensive approaches. (For a very thorough review of the approaches used to measure CSP see Wokutch and McKinney, 1991.) Among the most common approaches to defining CSP in terms of a single issue are CSP as pollution control (e.g. Bowman and Haire, 1975; Chen and Metcalf, 1980; Folger and Nutter, 1975; Spicer, 1975); as corporate crime (e.g. Wokutch and Spencer, 1987) and as corporate philanthropy (e.g. Galaskiewicz and Burt, 1991; Wang and Coffey, 1992). The more comprehensive approach has relied extensively on the use of the Fortune magazine corporate reputation survey (e.g. McGuire et al., 1988; Spencer and Taylor, 1987), survey data (e.g. Aupperle, 1984) or content analysis of annual corporate documents (e.g. Wolfe, 1991). Each of these approaches has attempted to look at the overall CSP construct in different ways.

Each of the previously cited approaches has brought important additions to our understanding of the theoretical and measurement issues of CSP. However, each has its own set of weaknesses. The single issue approaches limit themselves to one facet of what is commonly believed to be a multi-faceted construct. Those studies who attempt to develop a more broadly based approach also have their strengths and limits.

The studies using the Fortune magazine data have the advantage of the data being provided by presumably well informed managers outside of the focal firm. However as (Aupperle, 1991) and others have suggested there have been no attempts to validate these data and there is no discernible theory underlying the choice of variables. Finally, as other authors have noted (cf. Wartick, 1992) the “Responsibility to the community and the environment” item (the one most commonly used in CSP research) is highly correlated to the overall reputation score so it may not be a separate measure.

Survey based approaches such as Aupperle (1984) also have their own strengths and weaknesses. In Aupperle (1984), the survey instrument relied heavily on Carroll (1979) for its theory providing it with a sound basis upon which to build. The instrument also appears to employ many of Kerlinger’s (1986) and other’s methodological suggestions to improve the validity of the gathered data. In his own critique of his work (Aupperle, 1991) suggests that his data are limited by the use of single respondents in the firm and by some of the biases to which self-reported data are susceptible e.g. a social desirability bias. By modifying the Aupperle (1984) instrument, O’Neil, Saunders and McCarthy (1989) and Pinkston and Carroll (Forthcoming) were more effectively able to use the Aupperle scale.

Content analysis approaches (e.g. Wolfe, 1991) bring a different dimension to the investigation of CSP. By analyzing the written statements of corporations, a researcher is less susceptible to self-report bias. While corporate statements are, by definition, self-serving, proper content analysis methods can take this bias into consideration. Further, there is a large body of literature available (e.g. Weber, 1985) to help the researcher improve the rigor by which content analyses are done. However, the validity of content analyses rest on the coding schema adopted by the researcher. The more questionable the schema, the more the research is open to criticism.

In an effort to overcome many of the problems with the methods mentioned above, researchers have turned to a new source of data. Kinder, Lydenberg and Domini & Co. Inc., an investment advisory service specializing in “social choice” investing, has compiled CSP evaluations on approximately 800 publicly held firms. These data have been made available to researchers (e.g. Graves and Waddock, 1993; Ruf et al., 1993). The KLD ratings cover 9 areas of social performance including Community Relations, Employee Relations, Environment, Military Contracting, Nuclear Power, Product Liability, South African involvement, Women and Minority Issues and “other.” While no specific theory was used to develop this set of criteria,
they have become common social performance “screens” in the investment community (cf. Lydenberg et al., 1985). (Data that were precursors to the KLD ratings [Lydenberg et al., 1985] were used in Lerner and Fryxell’s [1988] study of the antecedents of CSP.) The KLD ratings hold promise in several areas for CSP researchers. First, the data cover several of the facets of the CSP construct. Secondly, the database currently covers about 800 firms whereas the Fortune database usually contains about 300 firms. Last, the data are evaluations done by individuals outside the focal firms so they are ostensibly more “objective” (or at least less susceptible to self-report biases) than data gathered via surveys or the content analysis of corporate documents.

Validity is the one major concern with the KLD ratings. Since the data have only recently been made available to researchers, no studies have appeared that have validated these data. The ratings represent, essentially, the collective beliefs of the KLD principals. The fact that the principals in KLD are well known and respected in the social investment field does give the ratings “face” validity. However, before they can be put extensively to use in empirical research, other elements of their validity have to be established.

Validating the KLD measures

Schwab (1980) leads one to believe that part of the reason why the general area of organization studies has failed to develop at a faster rate is that researchers have been lax in their examination of the validity of the measures that they employ. Consistent with Schwab’s (1980) arguments, this paper is an assessment of the validity of the KLD measures. Schwab (1980) suggests that there are three types of validity with which organization studies researchers should be concerned. The first he calls nomological (or theoretical) validity. In this approach to validation the researcher places the constructs within a theoretical network. Because the KLD measures already exist this approach to validity would be at best post hoc and hence is not a concern of this paper.

Schwab’s (1980) second type of validity, which he calls vertical validity, describes the extent to which a scale measures what it is supposed to measure, that is, effectively represents its latent construct i.e. construct validity. Our method for assessing construct validity is to do a version of it known as criterion validation. In this form of validation, one correlates the results of the measure in question with some other known (or at least accepted) method of measuring the same construct. An example of a relatively well known (or at least accepted) measure of CSP comes from the Fortune magazine corporate reputation survey data. If the KLD measures correlate well with the Fortune data we can have more confidence in these new measures. Without vertical (construct validity) researchers are in a weak position to use a measure for Schwab’s (1980) third type of validity that he calls horizontal (predictive validity).

Method

To assess the validity of the KLD ratings, this study correlated different combinations of the KLD ratings with three other sets of measures of overall CSP. The rest of this section details how the various scores were created.

Creating a KLD score. Creating a single score for the KLD data presented some unique challenges. A brief discussion of the KLD rating system is required to explain how the data were subsequently combined. A focal firm is given a “strength” and a “concern” score in each of the 9 categories. In each case, the score can be a “blank” indicated either insufficient information or no out-of-the-ordinary performance. The company can be given an “X” for a moderate concern/strength or “XX” for a strong concern/strength. Each non-zero evaluation is also classified by a “reason” which follows an essentially ordinal scale from lessor to greater importance. The first method for calculating a KLD score (KLD1) was simply to add all the ratings. A “bank” received a “0” while a strength X equaled 1 and a strength XX equaled 2. Concern scores were calculated as X = −1 and XX = −2. No attempt to use the “reasons” data was made at
this juncture because of the complexity involved in doing so. The data used in this study were the final 1991 ratings that were completed in August 1992.

However, this additive method implies that each category is equally important. Other researchers have shown that independent judges do not see the areas as equally important (e.g. Graves and Waddock, 1993; Ruf et al., 1993). Two additional KLD scores were created by multiplying the KLD ratings by the weights from the Graves and Waddock (1993), and the Ruf et al. (1993) studies. (Please see the respective papers for a detailed discussion of the respective rating systems.) It should be noted that since so few firms have a rating in the KLD “other” category, there is no weight listed for that area. The following table compares the ratings that the two studies developed.

To create KLD2, the Graves and Waddock (1993) weights were multiplied by the sum from the strength minus the concern rankings for each area except “other.” To create KLD3, the same process was used utilizing the Ruf et al. (1993) rankings. However, upon examination of both the Graves and Waddock (1993), and the Ruf et al. (1993) rankings, one can see a rather interesting break in the pattern. The ratings for nuclear power, military contracting and South African involvement are rated much lower than the other dimensions. The Ruf et al. (1993) data also suggest that this block was rated consistently lower than the others. Given that two sets of raters appear to have ranked these three criteria much lower than the other five, exploratory measures were created without them. Three additional scores (KLD4, KLD5 and KLD6) were created as described above but leaving out the scores for nuclear power, military contracting and South African involvement.

Other CSP measures. To assess the construct validity of the KLD measures, other measures of CSP had to be obtained. As mentioned previously, the Fortune corporate reputation survey data have been widely used as a measure of CSP. These data were used in several ways. First, the 1991 “Responsibility to the Community and the Environment” score for each firm was included in the data base (CSRB91). Second, as other authors have noted (cf. Wartick, 1992), the overall Fortune corporate reputation score is highly correlated with the “Responsibility to the community etc.” so the 1991 overall score was included also (OVER91). To see if there was any lagged relationship between the KLD scores and the Fortune data, the “Responsibility to the community etc.” and overall scores from 1989 and 1990 were also included in the data set. Finally, since the KLD ratings are not really point phenomena but rather summative over time, the three Fortune “Responsibility to the community etc.” scores and the three overall scores were averaged into composite scores (CSRB3Y and OVER3Y).

However, it could be argued that the Fortune data are not truly representative of CSP but rather simply the image that a particular firm has in the business community. To create an alternative CSP measure, data were collected from the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Relations</td>
<td>0.148</td>
<td>0.1250</td>
</tr>
<tr>
<td>Employee Relations</td>
<td>0.168</td>
<td>0.1828</td>
</tr>
<tr>
<td>Environment</td>
<td>0.142</td>
<td>0.1407</td>
</tr>
<tr>
<td>Military Contracting</td>
<td>0.086</td>
<td>0.0502</td>
</tr>
<tr>
<td>Nuclear Power</td>
<td>0.089</td>
<td>0.0740</td>
</tr>
<tr>
<td>Product Liability</td>
<td>0.154</td>
<td>0.2289</td>
</tr>
<tr>
<td>South African Involvement</td>
<td>0.076</td>
<td>0.0460</td>
</tr>
<tr>
<td>Women and Minority Issues</td>
<td>0.136</td>
<td>0.1523</td>
</tr>
</tbody>
</table>

Fig. 1. Weights from the Graves and Waddock (1993) and the Ruf et al. (1993) studies.
The Construct Validity of the KLD Social Performance Ratings Data

holdings list of the best known “social choice” mutual funds. The score on this variable was the number of times a firm was part of a fund portfolio (MATCHES). The idea here is that if a firm was chosen to be part of the portfolio it was a better than average firm in terms of CSP. There are two downsides to these data. First, there is a “floor” effect in that zero is the minimum value. Second, a zero score on this variable can indicate poor social performance or simply no data. With these limitations in mind, analyses were done that both included and excluded firms with a zero score. The data for this variable were collected in 1991 and 1992 from the latest available holdings lists from the top social choice mutual funds. From the information received from the funds, their turnover is very low so there is likely to have been little variation in their holdings lists over the period since the data were collected. The following table shows which funds were included in the list and the report used.

### Analysis and results

Pearson correlations were calculated between the variables of interest to determine the extent to which they were related. Table 1 presents the basic correlations between and among the variables. In this table, the six KLD scores are correlated with the 1991 *Fortune* “Responsibility to the community etc.” (CSR.B91) and overall reputation score (OVER 91) plus the number of MATCHES found in the mutual funds holdings lists. All correlation were significant at $p < 0.001$. The numbers of firms in each analyses are presented in parentheses.

The next step in the analysis was to see the effect of averaging the *Fortune* data over time. Table II presents the correlations with the six KLD scores and the three year averages for the *Fortune* “Responsibility to the community etc.” score (CSR.B3Y) and overall reputation score (OVER.3Y). All correlation were significant at

<table>
<thead>
<tr>
<th>Fund</th>
<th>Annual Report Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvert Social Investment Fund</td>
<td>5/31/90</td>
</tr>
<tr>
<td>Calvert-Ariel Appreciation Fund</td>
<td>7/1/90</td>
</tr>
<tr>
<td>Domini Social Index Fund</td>
<td>1/31/91</td>
</tr>
<tr>
<td>Dreyfus Third Century Fund</td>
<td>5/31/91</td>
</tr>
<tr>
<td>Parnassus Fund</td>
<td>11/30/90</td>
</tr>
<tr>
<td>PAX World Fund</td>
<td>6/30/90</td>
</tr>
<tr>
<td>Pioneer II</td>
<td>6/30/90</td>
</tr>
<tr>
<td>Pioneer III</td>
<td>6/30/90</td>
</tr>
<tr>
<td>CREF Social Choice</td>
<td>12/31/90</td>
</tr>
<tr>
<td>Working Assets Money Fund</td>
<td>6/30/90</td>
</tr>
</tbody>
</table>

Fig. 2. Social choice mutual funds from which data were collected.

| TABLE I  
Correlations with the calculated KLD scores and the criteria variables  

<table>
<thead>
<tr>
<th></th>
<th>KLD1</th>
<th>KLD2</th>
<th>KLD3</th>
<th>KLD4</th>
<th>KLD5</th>
<th>KLD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR.B91</td>
<td>0.30</td>
<td>0.35</td>
<td>0.36</td>
<td>0.40</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
</tr>
<tr>
<td>OVER.91</td>
<td>0.25</td>
<td>0.28</td>
<td>0.29</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
<td>(218)</td>
</tr>
<tr>
<td>MATCHES</td>
<td>0.30</td>
<td>0.31</td>
<td>0.30</td>
<td>0.29</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(776)</td>
<td>(776)</td>
<td>(776)</td>
<td>(776)</td>
<td>(776)</td>
<td>(776)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
TABLE II
Correlations between the calculated KLD scores and three average Fortune scores

<table>
<thead>
<tr>
<th></th>
<th>KLD1</th>
<th>KLD2</th>
<th>KLD3</th>
<th>KLD4</th>
<th>KLD5</th>
<th>KLD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSRB3Y</td>
<td>0.30</td>
<td>0.36</td>
<td>0.38</td>
<td>0.43</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>OVER3Y</td>
<td>0.23</td>
<td>0.28</td>
<td>0.28</td>
<td>0.31</td>
<td>0.31</td>
<td>0.30</td>
</tr>
</tbody>
</table>

$p < 0.001$ and there were 183 firms for whom all required data were present.

The next step in the analysis was to see how stable the relationships were over time. Table III presents correlations among the KLD scores and the 1989 to 1990 Fortune “Responsibility to the community etc.” (CSRB89 & CSRB90) and overall reputation scores (OVER89 & OVER90). All correlations were significant at $p < 0.001$.

The last analysis compared only those firms who had a non-zero holdings list score. Table IV presents the correlations among the six KLD scores, the average Fortune scores and the “matches” score. This analysis was done to determine if any possible “floor” effects were biasing the results. All correlations were significant at $p < 0.001$.

The final analysis looks at the interrelationships between the three most comprehensive criterion variables i.e. CSRB3Y, OVER3Y and MATCHES. Table V presents the correlation among these variables. A total of 183 firms had sufficient data to be included in this analysis (including zero scores on the Matches variable).

TABLE III
Correlations among the KLD and Fortune scores over time

<table>
<thead>
<tr>
<th></th>
<th>KLD1</th>
<th>KLD2</th>
<th>KLD3</th>
<th>KLD4</th>
<th>KLD5</th>
<th>KLD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSRP89</td>
<td>0.30</td>
<td>0.37</td>
<td>0.38</td>
<td>0.44</td>
<td>0.44</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
</tr>
<tr>
<td>CSRP90</td>
<td>0.27</td>
<td>0.33</td>
<td>0.36</td>
<td>0.40</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
</tr>
<tr>
<td>OVER89</td>
<td>0.24</td>
<td>0.30</td>
<td>0.30</td>
<td>0.33</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
<td>(208)</td>
</tr>
<tr>
<td>OVER90</td>
<td>0.16</td>
<td>0.21</td>
<td>0.22</td>
<td>0.25</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
<td>(211)</td>
</tr>
</tbody>
</table>

TABLE IV
Correlations among the KLD, Fortune and holdings list scores (restricted to those firms with a non-zero holdings list score)

<table>
<thead>
<tr>
<th></th>
<th>KLD1</th>
<th>KLD2</th>
<th>KLD3</th>
<th>KLD4</th>
<th>KLD5</th>
<th>KLD6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATCHES</td>
<td>0.18</td>
<td>0.20</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(499)</td>
<td>(499)</td>
<td>(499)</td>
<td>(499)</td>
<td>(499)</td>
<td>(499)</td>
</tr>
<tr>
<td>CSRB3Y</td>
<td>0.42</td>
<td>0.48</td>
<td>0.51</td>
<td>0.55</td>
<td>0.55</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
</tr>
<tr>
<td>OVER3Y</td>
<td>0.29</td>
<td>0.35</td>
<td>0.37</td>
<td>0.38</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
<td>(121)</td>
</tr>
</tbody>
</table>
TABLE V
Correlations between the Fortune three year average scores and the MATCHES score

<table>
<thead>
<tr>
<th></th>
<th>CSRB3Y</th>
<th>OVER3Y</th>
<th>MATCHES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSRB3Y</td>
<td>1.00</td>
<td>0.79</td>
<td>0.13</td>
</tr>
<tr>
<td>p</td>
<td>0.000</td>
<td>0.000</td>
<td>0.036</td>
</tr>
<tr>
<td>OVER3Y</td>
<td>0.79</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td>p</td>
<td>0.000</td>
<td>0.000</td>
<td>0.363</td>
</tr>
<tr>
<td>MATCHES</td>
<td>0.13</td>
<td>0.03</td>
<td>1.00</td>
</tr>
<tr>
<td>p</td>
<td>0.036</td>
<td>0.363</td>
<td>0.000</td>
</tr>
</tbody>
</table>

TABLE VI
Correlations between the Fortune three year average scores and the MATCHES score (firms with non-zero MATCHES scores only)

<table>
<thead>
<tr>
<th></th>
<th>CSRB3Y</th>
<th>OVER3Y</th>
<th>MATCHES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSRB3Y</td>
<td>1.00</td>
<td>0.81</td>
<td>0.15</td>
</tr>
<tr>
<td>p</td>
<td>0.000</td>
<td>0.000</td>
<td>0.047</td>
</tr>
<tr>
<td>OVER3Y</td>
<td>0.81</td>
<td>1.00</td>
<td>0.06</td>
</tr>
<tr>
<td>p</td>
<td>0.000</td>
<td>0.000</td>
<td>0.246</td>
</tr>
<tr>
<td>MATCHES</td>
<td>0.15</td>
<td>0.06</td>
<td>1.00</td>
</tr>
<tr>
<td>p</td>
<td>0.047</td>
<td>0.246</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table VI presents the same correlation with the 121 who had Fortune scores and a non-zero MATCHES score.

Discussion

The assessment of validity is not a binary phenomenon. It is not a situation where a measure is either valid or it’s not. Rather, since validity is a continuous phenomenon, we can only say the degree to which a measure is valid. Theoretically, a measure could be perfectly valid (i.e. receive some sort of validity coefficient of 1.00). However, in practice, validity scores are usually less than 1.00. Since perfection appears to be highly unlikely, the researcher is left with the task of ascertaining “how much is enough.”

The basic question in this research is – do the KLD ratings correlate sufficiently with other measures of corporate social performance. These data suggest that the answer to that question is a qualified “yes.” As one can see from the data presented above, the KLD scores correlate with other criterion variables with values ranging from 0.18 to 0.55. The data from Table I suggests that the Fortune “Responsibility to the community etc.” score is the most highly correlated with the various KLD measures. We also see from Table I that weighing the KLD scores generates better correlations than a simple summation. Further we see that dropping the nuclear power, military contracting and South African involvement ratings and creating a weighted score increases the correlation even more with CSR91 but appears to have little effect on the correlations with either the holdings list variable or the overall Fortune reputation score.

The data in Table II help us explore the extent to which using scores averaged over time improved the correlations. In the case of the correlations between CSRB3Y and the various KLD measures, we do see a slight increase in the correlations – especially for the KLD scores where the nuclear power, military contracting and South African involvement ratings had been dropped. We see very little difference between the table for the correlations among the KLD scores, the overall Fortune score (OVER3Y) and the holdings list score (MATCHES).

Table III helps us examine the stability of the relationships among the variables over time. A rather interesting phenomenon is observed by looking at this table. One might assume that the further back in time the Fortune data went, the less powerful the relationship might be between them and the KLD scores. We observe a different pattern in the data. The relationships among the variables were stronger for the 1989 Fortune data than for the 1990. Interestingly, the relationships with the 1991 Fortune data returned to the levels observed with the 1989 data.

Table IV allows us to examine any floor effect that might be present with the holdings list data. By only examining the correlations for those firms with a non-zero holdings list score, we get a very interesting picture. The relationship between the KLD measures an the holdings list variables declines with the inclusion of only the
non-zero firms. Further research is necessary to
determine if the presence of the zero scores
falsely inflated the relationship or allows for a
more accurate description. We also see a very
large jump in the relationship of the KLD scores
and the Fortune data. Since we are only exam-
ing firms who have been selected into the
holdings lists, their actual CSP is likely to be
higher. As such, the relationship between the
KLD scores and the perceptions of the Fortune
respondents are likely to be higher.

Tables V and VI suggest that even though the
KLD scores are related to the criterion scores,
the Fortune data and the holdings lists scores may
be tapping different parts of the CSP construct.
We see that the CSRBJ3Y and the MATCHES
variables are correlated but at a relatively low
r = 0.13. This statistic improves slightly to r =
0.15 when we look at those firms who had a
non-zero matches score. The holdings list
variable is not significantly related to the overall
average Fortune reputation score.

We see from these results that the KLD ratings
seem to be capturing at least part of the same
construct as do the other measures of corporate
social performance and hence can be considered,
to that extent valid. The data from Tables V and
VI also suggest that the KLD rankings are cap-
turing different parts of the CSP construct since
the Fortune data and the MATCHES variable are
correlated, but only at a relatively low level.

However, since as mentioned previously,
validity is a continuous rather than a binary
concept, the question becomes – are the KLD
scores valid enough? Without a theoretical net
within which to embed the scales, we can’t use
Schwab’s (1980) nomological criterion. We are
faced with the task of deciding if the correlations
presented in these pages are sufficient. To do so
requires an additional look at the other criterion
variables we examined.

The MATCHES variable was created by
examining the holdings lists of “social choice”
mutual funds. These funds are run by professional
money managers who are skilled in social eval-
uation and presumably can evaluate the social
performance of a focal firm. However, the choice
for each mutual fund is a binary one – either
the firm is included in the portfolio or not. Even
though this present research was able to create a
continuous variable by looking at the number of
funds that held a focal firm’s stock, this measure
probably does not have as much sensitivity as
do the KLD measures. Further, although as up
to date as were possible, there was a lag time
between the holdings lists data and the data of
the KLD assessments. While the social choice
mutual funds do not “churn” their portfolios
often, the variation in dates may have had some
effect. As such, while this variable was helpful
in examining the validity of the KLD ratings, it
may not be the best choice for a final standard.

The various Fortune data may be of more
benefit to our analysis do to their nature. Both
the overall score and the “Responsibility to the
community etc.” scores are continuous by nature
and hence more similar to the calculated KLD
scores. Further, the panels of industry executives
are made up of people who should be relatively
familiar with the activities of peer firms. (The
Fortune data have been criticized, however,
by participating executives because they do not
believe the previous statement to be true.)
Regardless of whether the executive are or are
not an informed panel, we see the KLD scores
correlate with the Fortune data with “r” values
ranging from 0.30 to 0.55 depending on the data
set. We also see that there is relatively little lagged
effect because the correlations stay similar over
time.

It should be noted that the modified scales that
were constructed without the nuclear power,
military contracting and South African involve-
ment ratings correlate better with the Fortune
data than do the scores from the 8 topic-based
ratings. This could indicate two things: (1) The
3 left-out issues really aren’t very important to
any one but social activists and the Fortune raters
don’t take these issues into consideration when
making their choices or; (2) At least in the case
of South African evolvement, the dismantling of
apartheid has rendered this dimension irrelevant.
Further research could also help clarify this issue.
It should also be noted that the highest correla-
tions occurred when the data set was reduced to
only the firms that had both appeared in the
Fortune dataset and had non-zero MATCHES
scores. This suggests that the KLD scores may
be better in evaluating "saints" rather than "sinners" (cf. Wokutch and Spencer, 1987) i.e. high social performers rather than low or poor performers.

Future research

While the analyses in this paper have established that the KLD scores are tapping into the same construct as do the other measures, the correlations were not overwhelming. In every case, the correlations explained less than 50% of the variance in either variable. The discussion above suggests that some of the problem may be weaknesses in the criterion measures. To remedy this problem, further research is needed to compare the KLD scores with additional CSP measures. Two other measures that are available for comparison are corporate philanthropy data and toxic release survey data. Both sets of data have been used in the past as measures of CSP. If these variables are also correlated to the KLD scores, we can develop even more confidence that the KLD scores are tapping into the CSP construct.

Acknowledgement

I wish to acknowledge the assistance of several individuals and groups without whose help this paper would not have happened. First, thanks must go to Kinder, Lydenberg & Domini and especially to Steve Lydenberg for providing access to and much assistance with their data. Thanks also go to Fortune magazine for providing the Corporate Reputation Survey data. My thanks also go to Karen Paul, Sandra Waddock and their respective co-authors for providing their weights for the KLD scores and their comments on this project. Finally, many thanks must go to Jef Shilling for his data collection and computer machinations.

Conclusion

As the field of Social Issues Management attempts to develop a common body of knowledge, measures and data, the assessment of the psychometric properties of said measures and data becomes critical. The results presented in this paper suggest the KLD social performance ratings data are measuring at least part of the same CSP construct as to the Fortune magazine data and a measure derived from the holdings lists of a large number of social choice mutual funds. However, while the correlations were substantial they were not overwhelming. As noted above, the moderate level of correlations could stem from several sources. In fact, the level of the correlations could stem from the fact that the KLD data could be a better measure of CSP than either of the other two. In measurement terms, the error variance could stem from the other measures rather than those derived from the KLD data. Future research has been proposed that could help clarify this point. In any case, researchers interested in studying corporate social performance now can have confidence in the KLD measures and feel secure in the idea that the this new data does tap into the core of the social performance construct.

Bibliography

Chen, K. H. and R. W. Metcalf: 1980, 'The Relationship between Pollution Control Record


*College of Business Administration, Division of Management, University of Oklahoma, Norman, OK 73019-0450, U.S.A.*