ON THE VALIDITY OF SUBJECTIVE MEASURES OF COMPANY PERFORMANCE

TOBY D. WALL
Institute of Work Psychology
University of Sheffield, U.K.

JONATHAN MICHIE
School of Management and Organizational Psychology
University of London, Birkbeck, U.K.

MALCOLM PATTERSON
Institute of Work Psychology
University of Sheffield, U.K.

STEPHEN J. WOOD
Institute of Work Psychology
University of Sheffield, U.K.

MAURA SHEEHAN
Graduate School of Management
University of Dallas

CHRIS W. CLEGG
Institute of Work Psychology
University of Sheffield, U.K.

MICHAEL WEST
Business School
Aston University, U.K.

Subjective measures of company performance are widely used in research and typically are interpreted as equivalent to objective measures. Yet, the assumption of equivalence is open to challenge. We compared the use of both types of measure in 3 separate samples. Findings were consistent in showing that: (a) subjective and objective measures of company performance were positively associated (convergent validity); (b) those relationships were stronger than those between measures of differing aspects of performance using the same method (discriminant validity); and (c) the relationships of subjective and objective company performance measures with a range of independent variables were equivalent (construct validity).

We acknowledge support from the Economic and Social Research Council (UK), the Leverhulme Trust, the Royal Economic Society, the University of London Central Research Fund, and the University of Dallas, who funded the studies on which this paper is based. Thanks go also to the colleagues who contributed to the research, including P. Frangou, R. Lawthom, E. McDonald, S. Milner, K. Pepper, and C. Stride.

Correspondence and requests for reprints should be addressed to Toby D. Wall, Director, Institute of Work Psychology, University of Sheffield, Sheffield, S10 2TN, UK; t.d.wall@Sheffield.ac.uk

COPYRIGHT © 2004 PERSONNEL PSYCHOLOGY, INC.
The measurement of company performance is of considerable importance in many areas of personnel psychology and management research, as investigators examine whether organizational practices and processes affect "the bottom line." Studies of human resource management have focused on how practices, such as selection, training, empowerment, and communications, collectively contribute to company performance (e.g., Becker & Huselid, 1998; Huselid, 1995). Similarly, in the strategy and marketing literature, there have been many attempts to link strategy to performance (Dess & Robinson, 1984; Pleshko & Souiden, 2002; Robinson & Pearce, 1988; Wensley, 1999). In operations management, interest has been on the performance benefits for the company of practices such as just-in-time and total quality management (e.g., Cua, McKone & Shroeder, 2001; Sakakibara, Flynn, Schroeder & Morris, 1997; Shah & Ward, 2003). In addition, within industrial relations and economics, studies have examined the relationship of union representation and employment practices (e.g., payment systems) with company and workplace performance (e.g., Fernie & Metcalfe, 1995; Freeman & Medoff, 1984; Metcalfe, 2003; Patilbanda & Chandra, 1998).

Although some studies use objective measures of company performance, such as productivity, profit, or return on assets, typically taken from externally recorded and audited accounts, many rely on subjective measures as reported by respondents. For example, approximately half of the published studies on HRM and performance use such subjective data (e.g., Delaney & Huselid, 1996; Guthrie, 2001; Wright, McCormick, Sherman & McMahan, 1999; Youndt, Snell, Dean & LePak, 1996); and the large majority of the industrial relations studies (notably those from the U.K.'s Workplace Employee Relations Survey) have also been based on subjective measures (e.g., Addison & Belfield, 2001; Wood & de Menezes, 1998).

There are good reasons why subjective measures of company performance have been and will continue to be employed. One is that they are cost effective because such performance data can be collected through questionnaire or interview surveys that simultaneously elicit information on practices. The more fundamental reason, however, is that for certain types of organization and levels of analysis there may be no viable alternative. For many public service and voluntary sector organizations, and indeed for many smaller enterprises, there are no appropriate financial records. Even for organizations with such records, the data may be aggregated in a way that is not compatible with the level of analysis for the HRM or employment relations practices of interest. In many companies, such practices differ considerably across workplaces or sites, whereas the available financial performance data are for the enterprise as a whole.
The key point for our present purposes, however, is that in interpreting research findings, subjective and objective measures of company performance have been treated as equivalent. This is illustrated in reviews of research on human resource management and company performance (Gerhart, Wright, & McMahan, 2000; Guest, 1997; Wood, 1999). Moreover, it parallels the approach to subjective and objective measures of individual employee performance, where "researchers continue to treat different performance measures synonymously... conclusions are intended to generalize to a broad performance construct, irrespective of the measure(s)" (Bommer, Johnson, Rich, Podsakoff, & Mackenzie, 1995, p. 588).

In considering this assumption, it is important to recognize that, in practice, the equivalence of subjective and objective measures has been potentially confounded by two other differences in the measures used. First, subjective measures have tended to focus on overall performance, whereas objective measurement has typically used more specific financial indicators. In this respect, subjective measures of overall performance have either been overall performance ratings, where performance has been largely undefined (e.g., Dess & Robinson, 1984), or have consisted of a set of more specific items that have been aggregated to create a composite score (e.g., Hoque, 1999; Wright et al., 1999; Youndt et al., 1996). For example, Dess and Robinson (1988) used a single item, simply asking respondents to rate "overall firm performance/success" (p. 268); whereas Delaney and Huselid (1996) combined responses to four items rating performance with respect to marketing, growth in sales, profitability, and market share (p. 956). In contrast, those deploying objective measures typically have used two or three more specific indices from among the many possible, and presented results for each separately. For instance, Guest, Michie, Conway, and Sheehan (2003) used productivity and profit per employee, and Huselid (1995), productivity, gross return on assets, and Tobin's $Q$.

The other difference is that subjective measures have tended to ask respondents to rate their company's performance relative to their competitors, whereas objective measures have been absolute (e.g., $ profit per employee). We can illustrate this relative aspect of subjective measures by expanding on the examples given above. Dess and Robinson's (1988) single-item measure required respondents to compare their company to "firms of similar sales volume in your industry and region" (p. 268); and Delaney and Huselid's (1996) 4-item measure asked "compared with other organizations that do the same kind of work, how would you compare the organization's performance over the last 3 years" (p. 956).
Thus, when subjective measures are treated as equivalent to objective ones, this glosses over the fact that there are other differences in measurement both within subjective and objective measures, and especially between them. Nonetheless, there are three reasons for believing that this confounding may not be as big a problem as it first appears. First, subjective measures are directed at employees of CEO, director, or equivalent level, for whom financial considerations of the kind captured by objective measures are likely to dominate their view of company performance. So, even general subjective measures are likely to reflect the more specific objective ones. Second, evidence exists that subjective measures of overall performance are closely associated with more specific subjective measures. In Dess and Robinson’s (1988) study, the single-item overall performance rating correlated strongly with reports of total sales \( r = .80 \) and with return on assets \( r = .79 \). Robinson and Pearce (1999) describe similar findings. Third, Bommer et al.’s (1995) meta-analysis of 50 studies, in the parallel domain of employee performance, found no effect of such differences in measurement approach (i.e., overall \( v \) composite, and relative \( v \) absolute), in the extent to which subjective measures were associated with objective ones.

For the purposes of this paper, therefore, our working assumption is that such differences in specificity and relativity between subjective and objective measures do not preclude sensibly investigating the relationship between the two. Indeed, given that such differences are a feature of work to date, it is important that they are represented in any investigation. Reflecting this, in the two studies reported in this paper, one uses a single-item overall subjective rating of relative performance alongside a specific objective measure (profit), whereas the other study uses relative subjective performance ratings at the same level of specificity as for the absolute objective measures (productivity and profit in both cases). Clearly, if such studies yield similar findings for the equivalence of subjective and objective measures, it would reinforce the view that content issues were of minor significance. Alternatively, if they provide differential support for equivalence, this would suggest ways of improving subjective measures (and possibly also objective ones).

Our primary concern, however, is to address the threat to validity that comes from potential error in subjective measures. Here, two types of error are especially significant. First, if subjective performance measures contain random error, for example as a result of respondents remembering figures incorrectly, guessing, or confusing the accounting period of interest with an earlier or later one, then the effect will be to attenuate any real underlying relationship with associated variables of interest (Type II error, or false negatives). More troublesome is the possibility of systematic bias creating relationships between practices and
performance that do not really exist (Type I error, or false positives). Such "common-method variance" bias is a danger where information on the practices and performance is obtained from the same respondent, a common procedure in the literature. It can arise for a variety of reasons, such as a general tendency on the part of individuals to be positive or negative in their outlook, to use higher or lower points on response scales, or for their ratings of practices to be colored by their assessments of financial performance.

Recent research has begun to address the question of the validity of subjective measures in the HRM area of inquiry but has focused on the measurement of practices rather than performance. This is exemplified by a recent series of articles (e.g., Gerhart et al., 2000; Gerhart, Wright, McMahan, & Snell, 2000; Huselid & Becker, 2000) on questionnaire measures of HRM practices. Rather less attention has been paid to the validity of measures of subjective performance.

Two empirical studies present evidence relating to the question of the validity of subjective company performance measures. One, by Baer and Frese (2003), involved a study of 47 medium-sized German industrial companies (100 to 900 employees). Subjective performance was measured by a questionnaire completed by a managerial representative from either the personnel or marketing department. There were two items: "How successful is your organization in comparison to other companies in the same line of industry and of about the same size?" and "To what degree has your organization achieved its most important goals?" The items were combined into a single index that had an internal consistency (Cronbach's alpha) of .83. The objective indicator of performance was an accountancy-based index of return on assets, calculated by dividing annual profit by invested capital. The product-moment correlation between the subjective and accountancy-based performance measures was .41.

The second study is that described by Guthrie (2001). The sample comprised senior management respondents (e.g., HRM director, managing director) from 164 New Zealand companies employing 100 or more people. Subjective performance (productivity) was operationalized as reported sales per employee, calculated from responses to two questionnaire items asking for the most recent estimates of annual sales and total number of employees. The directly corresponding objective financial data (i.e., sales per employee) were obtained for a subsample of 65 companies. The product-moment correlation between the two measures was .81.

The above studies provide some evidence of convergent validity in that the measures of subjective performance were associated with corresponding objective measures. To this we can add Machin and Stewart's
(1996) finding relating to predictive validity, that a subjective measure of financial performance predicted the subsequent closure of companies several years later. However, three small-scale studies do not provide a strong foundation from which to generalize. Moreover, because this was not their prime purpose, none of the studies address the validity of subjective measures in sufficient depth. We need to expand the analysis of validity in two ways. First, following Campbell and Fiske's (1959) "multi-trait-multimethod matrix" rationale, validity requires not only that subjective and objective measures of the same performance construct are associated (i.e., convergent validity), but also that their association should be stronger than that between measures of related but different constructs using the same method (i.e., discriminant validity). For example, the relationship between subjective and objective measures of productivity, and between subjective and objective measures of profit, should be stronger than the relationships between subjective measures of productivity and profit, or between objective measures of productivity and profit. In other words, different ways of measuring the same construct should show stronger correspondence than the same way of measuring different constructs.

The second way of adding to our understanding of subjective performance measures is to examine a key aspect of their construct validity. Even for a subjective performance measure with good convergent validity, the amount of unique variance in that measure leaves plenty of scope for relationships to be found with other variables simply as a result of common-method variance, when they would not be found were an objective measure used. It is thus important to determine not only whether subjective and objective measures of performance are associated with one another, but also whether they are associated with other variables of interest in the same way. The question this answers is: Does the use of a measure of subjective performance lead to findings that differ from those based on objective measures of performance? If there is no such difference, we can be more confident that the subjective performance does not lead to erroneous conclusions. Given evidence of convergent and discriminant validity, this final stage is the acid test.

This paper addresses these issues. It capitalizes on two recent studies (one with two separate samples) in which both subjective and objective performance were measured together with organizational practices. Thus, the paper provides additional evidence on the convergent validity of subjective measures of company performance, and new evidence on their discriminant and construct validity. In both studies, we use objective performance as the criterion against which to judge the validity of subjective performance, making the working assumption that such objective measures are themselves valid. In practice, however, it is clear
that objective measures are not free from error, and we consider the implications of this in the discussion.

Study 1: Single-Site Manufacturing Companies

The primary aim of this study was to examine the relationship between the use of selected management practices (just-in-time, total quality management, job enrichment, and skills enhancement) and subsequent objective company performance (productivity and profit). Some of these findings have been described elsewhere (Patterson, West, Lawthom, & Nickell, 1997; Patterson, West, & Wall, in press). This paper focuses on as-yet unreported data on performance concurrent with the measurement of the practices, for which both subjective and objective measures were available.

Method

Sample and procedure. The sample comprised 80 U.K. manufacturing companies employing from 60 to 1,150 employees. A key criterion in selecting the companies was that they should be single site so that the unit of analysis for the practices would correspond directly to that for financial performance.

Data on management practices were collected through onsite, semi-structured interviews with senior managers and directors. The interviews were conducted with the person primarily responsible for the practice in question (e.g., the quality manager for total quality management; the HRM manager for skills enhancement). On average, three different managers or directors were interviewed per company. Evidence from each interviewee was cross checked with that from the others, and with additional information from company documents and a tour of the manufacturing facility. Based on this information, two researchers, unaware of company performance, independently rated the extent of use of each of the practices. Subjective performance was obtained using a rating scale completed by the most senior interviewee, and objective performance was determined subsequently from audited accounts.

Measures. To measure subjective financial performance the interviewee was asked “What is your company’s performance in comparison to your main competitors.” Ratings, on a 5-point response scale running from 1 = much worse to 5 = much better, were obtained both for the year in which the interviews took place ($M = 3.78; SD = .87$), and for the previous 5 years as a whole ($M = 3.38; SD = 1.15$).

Objective performance was measured using financial data extracted by economists from audited records lodged in Companies’ House. Profit was calculated as the logarithm of the financial value of sales less costs
before profits tax. In order to control for changes due to movements in prices, profit was deflated by the producer price index for the sector in which the company belonged. The data were obtained both for the year in which the study took place ($M = 3.60; SD = 5.49$), and for the previous 3 years ($M = 2.51; SD = 5.61$). We could not go back as planned for the full 5 years covered by the subjective measure because of unanticipated limitations in the financial data. However, because the measures cover 3 years in common, together with the fact that year-on-year profits are typically strongly correlated (e.g., Guest et al., 2003; Patterson et al., in press), we are confident that the measures correspond sufficiently for present purposes.

The use of two raters to measure each of the management practices enabled us to determine the level of interrater agreement. As there is some controversy over the most appropriate way of assessing this (Kozlowski & Hattrup, 1992), we calculated two contrasting but widely used indices. One was the intraclass correlation coefficient (ICC), of which there are two main forms: ICC2 that indicates whether organizations can be reliably differentiated from one another in terms of multiple ratings on the variable of interest (James, 1982); and ICC1 that shows the extent to which two or more individuals agree in their ratings (Shrout & Fleiss, 1979). The latter was the appropriate one for current purposes. The second index was Spearman's rank-order correlation ($\rho$) between the two sets of ratings.

Four management practices were included in the study. The measure of just-in-time assessed the extent to which the company used procedures to minimize inventories and lead-times, and was rated on a 5-point scale ($M = 2.95, SD = .63$). The ICC1 was .80 and $\rho$ was .88. The measure of total quality management was based on: the existence of procedures to ensure the quality of supplies; use of methods of quality assurance applied to production processes and end-products; and the use of quality techniques (e.g., statistical process control, quality circles) and training in quality. Ratings were on a five-point scale ($M = 2.53, SD = .96$). The ICC1 was .96 and $\rho = .97$. The measure of job enrichment covered the amount of responsibility given to shop-floor employees, including delegation of management of materials and work in progress, degree of multi-skilling and job rotation time. Ratings were on a 7-point scale ($M = 3.79, SD = 1.08$). The ICC1 and $\rho$ were both .97. Finally, the index of skill enhancement captured the extent to which the company emphasized skill development for shop-floor employees. It included the extent of systematic selection methods, training needs assessment, and training time. Ratings were on a five-point scale ($M = 2.76, SD = .69$). The ICC1 was .98 and $\rho = .97$. Further detail of the measures can be found in Patterson et al. (2003).
TABLE 1
Product–Moment Correlations Among Current and Previous Subjective and Objective Company Performance Measures in a Sample of 80 U.K. Single-Site Manufacturing Companies

<table>
<thead>
<tr>
<th></th>
<th>Subjective performance</th>
<th></th>
<th>Pre</th>
<th>Previous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current</td>
<td>.40**</td>
<td>.27*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Previous</td>
<td>.26</td>
<td>.42**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: N varies from 70–80 depending on missing values.
*p < .01    **p < .001

TABLE 2

<table>
<thead>
<tr>
<th>Management practice</th>
<th>Current performance</th>
<th></th>
<th>Pre</th>
<th>Previous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subjective</td>
<td>Objective</td>
<td>Subjective</td>
<td>Objective</td>
<td></td>
</tr>
<tr>
<td>Just-in-time</td>
<td>-.01</td>
<td>.07</td>
<td>-.08</td>
<td>-.03</td>
<td></td>
</tr>
<tr>
<td>Total quality management</td>
<td>.04</td>
<td>.13</td>
<td>.03</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Job enrichment</td>
<td>.31*</td>
<td>.41*</td>
<td>.38**</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Skills enhancement</td>
<td>.06</td>
<td>.18</td>
<td>.06</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>

*Note: N varies from 70–80 depending on missing values.
*p < .01    **p < .001

Results

First, we examined the relationship between the subjective and objective measures of performance to assess convergent and discriminant validity. The product-moment correlations between the two kinds of measure, for both current and previous profit, are shown in Table 1. It is clear that the subjective and objective measures are statistically significantly related, in that the correlation between the two for the current year is .40 (p < .001) and for previous years is .42 (p < .001). Thus, the measures show a degree of convergent validity. At the same time, the weaker off-diagonal correlations, between current subjective performance and previous objective performance (.26, n.s) and between previous subjective performance and current objective performance (.27, p < .01), are lower, and, hence, consistent with the requirements for discriminant validity. Nonetheless, the latter finding should be treated with caution as tests for the convergent coefficients being greater than the off-diagonal ones (Hotelling’s t-test for the difference between correlated coefficients of correlation) were not statistically significant in any of the four cases, ranging from p < .08 to p < .15.
The question of construct validity was approached by determining whether the relationships of the subjective performance measure with other variables is equivalent to that of the objective measure. Table 2 presents the product-moment correlations for each type of measure for both current and previous performance with measures of the four management practices. The pattern of findings shows that the subjective measures relate to other variables in a way that corresponds closely to that of the objective measures. Indeed, tests of the difference between the correlations of objective and subjective performance with each practice show that in no instance are they statistically significantly different. Even if one were to interpret the findings using conventional statistical significance levels, as might occur in a substantive study, the conclusions using the subjective and objective measures would be essentially identical. In every case where the subjective measure has a statistically insignificant relationship with a management practice, so too does the objective measure. On the other hand, where the subjective performance measures show a statistically significant relationship with a management practice (i.e., job enrichment), the objective measures follow the same pattern. The only anomaly is the failure of the objective previous performance measure to show a statistically significant relationship with job enrichment (.22, n.s) where the subjective performance measure does (.38, p < .01), but the former coefficient is nevertheless larger than for any other practice and, as noted above, the difference between the two is not statistically significant.

Discussion

The findings from this study suggest considerable validity of the subjective performance measure judged against the objective one. With respect to convergent validity, the association between subjective and objective measures for the equivalent period is evident, though relatively modest. Discriminant validity is also apparent because the associations between subjective and objective measures for different periods are weaker than those for equivalent periods. Most importantly, there is clear evidence of construct validity, given little indication that the subjective measure would lead to different conclusions concerning relationships with other variables than would tests using the objective measures. Notwithstanding the one anomaly, the results overall do not support the view that the subjective measure leads to overestimation of relationships with other variables relative to the objective measure. Bear in mind, however, that the opportunity for such overestimation as a result, say, of common-method variance bias, is restricted in this study, as the measures of the practices and subjective performance came from independent sources.
Nonetheless, there are clear limitations to this initial study. First, it is based on a relatively small sample. Second, though it was possible to address discriminant validity using different periods of performance (current and previous), the previous periods covered for the subjective and objective measures were not identical and there was no test involving different aspects of performance (e.g., profit vs. productivity). Finally, with regard to construct validity, the findings are based on only three associated variables, only one of which showed a positive relationship. Thus, the opportunity to register negative results was limited.

Study 2: Performance and Management Practices in Single- and Multi-Site Companies

The second study spans and adds to the previous one by including both single- and multi-site companies, measuring both productivity and profit for the current and previous years, using directly equivalent subjective and objective performance measures, and covering additional management practices. Moreover, it encompasses both the manufacturing and service sectors. Substantive findings from the same data set are reported in Michie and Sheehan (2001, 2003). The analyses here, however, are unique to this paper.

Method

Sample and procedure. The study is based on a stratified sample of publicly quoted U.K. manufacturing and service sector firms with more than 50 employees. Two dimensions were used to stratify the sample: organization size and the primary sector of business activity. Five sizes of company were specified for sampling purposes, those employing 50–99, 100–499, 500–999, 1,000–9,999, and 10,000 or more employees. Fourteen sectors were identified (9 in manufacturing and 5 in services) using the 1992 U.K. Standard Industrial Classification (SIC) codes. The EXTEL database was used to locate all organizations with the above characteristics, which resulted in a total of 1,280. This database is updated weekly and features high-value, structured financial data, including detailed coverage of more than 14,000 quoted companies from 53 countries with up to 15 years of history. Data include balance sheets, profit and loss statements, share information, and other financial details. Holding companies were removed, which reduced the sample to 934. The aim was to obtain a total quota of 33% of the “population” and to achieve representation within the stratified sample cells.

A market research company was commissioned to conduct a telephone survey of the selected companies. The following protocol was followed: (a) The name and fax number of the senior manager with
HRM responsibilities was confirmed by telephone. (b) This person was faxed a one page briefing on the objectives of the survey and the expected length of the interview (around 30 minutes). Confidentiality of survey participants was emphasized. (c) The day after faxing, the person was contacted by telephone to see if they would consent to the telephone interview and, if agreeable, a date and time for the interview was set. Of the 934 potential respondents identified, 559 declined, and 19 agreed but subsequently failed to complete the interview. Thus, 369 interviews were completed successfully.

In 78% of cases, the telephone interviews were with the director of human resources, personnel, or employee relations of the company. Where this person was not available, an alternative senior colleague was interviewed, typically the chairperson, managing director, chief executive, manufacturing director or production director. In only 73 instances was the interviewee unable to answer all questions, with nonresponse typically involving matters about performance. In such cases, the name of an additional appropriate person was solicited and that person was contacted to complete the missing parts of the interview. In total, 369 interviews were completed, of which, because of missing data, 361 were usable. Sixty-seven percent (242) of respondents were from manufacturing companies and 33% (119) were in service sector establishments. Twenty-three percent of the respondents (83) were from single-site organizations and 73% (278) from multi-site ones.

*Measures.* Subjective performance was measured in the interviews in two ways. One question was “How does labor productivity at this site compare with other establishments in the same industry?”; and the other “How does this site’s financial performance, that is, profitability, compare with other establishments in the same industry?” Responses were obtained on a 5-point response scale running from 1 = *much worse* to 5 = *much better*, both for the year in which the interviews took place, and for the previous 3 years. The means and standard deviations for the single-site organizations were: $M = 4.10$, $SD = .48$ and $M = 3.52$, $SD = .55$ for current productivity and profit respectively; and $M = 3.63$, $SD = .55$ and $M = 3.11$, $SD = .61$ for previous productivity and profit respectively. The corresponding figures for multi-site organizations were: $M = 4.36$, $SD = .60$ and $M = 3.81$, $SD = .63$, for current productivity and profit; and $M = 4.19$, $SD = .69$ and $M = 3.52$, $SD = .74$, for previous productivity and profit.

Objective performance was measured using financial data extracted from the EXTEL database. Productivity was calculated as gross sales per employee. Profit was calculated as the pretax financial value of sales less costs per employee. Consistent with the subjective performance measures, both objective performance indicators were calculated for the year

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
in which the study took place, and also for the previous 3 years. To control for price movements, the two variables were deflated by the producer price index for the sector in which the company belonged. To deal with the nonnormal distribution of the data, we used the logarithm of the raw scores. The means and standard deviations for the single-site organizations were: $M = 8.13, SD = 1.77$ and $M = 5.63, SD = 1.98$ for current productivity and profit, respectively; and $M = 7.06, SD = 2.13$ and $M = 5.09, SD = 2.37$ for previous productivity and profit, respectively. The corresponding figures for multi-site organizations were: $M = 9.94, SD = 2.03$ and $M = 6.22, SD = 2.29$ for current productivity and profit; and $M = 8.31, SD = 2.57$ and $M = 5.48, SD = 2.60$ for previous productivity and profit.

We included nine management practices in the survey, four equivalent to those in Study 1, namely, just-in-time, total quality management, empowerment, and skills enhancement, with the additional five being supply-chain partnering, business process reengineering, team-based working, outsourcing, and total productive maintenance (for manufacturing companies only). All were measured using the procedure adopted by Waterson et al. (1999) of providing the interviewee with a definition of the practice and obtaining a rating of use. Respondents described the extent of use of each practice on the following five-point scale: not at all, a little, moderately, a lot, and entirely. For multi-site enterprises, respondents were asked to respond for the most typical site. Means and standard deviations are as follows: just-in-time, $M = 3.23, SD = .43$; total quality management $M = 3.85, SD = .60$; empowerment $M = 3.89, SD = .42$; skills enhancement $M = 4.11, SD = .38$; supply-chain partnering $M = 3.76, SD = .73$; business process reengineering $M = 2.98, SD = .81$; team-based work $M = 3.53, SD = .61$; outsourcing $M = 3.90, SD = .68$; total productive maintenance $M = 3.16, SD = .37$.

Results

Even after logarithmic transformations the data were not normally distributed, so we used an appropriate nonparametric method of analysis, the Spearman rank-order correlation. We conducted the analyses for the sample as a whole, by type of site (single and multiple), and by sector (manufacturing and services). The findings for manufacturing and services were equivalent so we do not differentiate between them here (that breakdown is available on request from the second author). There were, however, minor differences in the pattern of results for the single- and multi-site companies, and we present the results accordingly. This provides an additional point of comparison because for the single-site companies the measures of the practices and performance correspond to the same unit of analysis (the company as a whole), as in Study 1;
<table>
<thead>
<tr>
<th>Objective performance</th>
<th>Single-site companies (n = 83)</th>
<th></th>
<th>Multi-site companies (n = 278)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current Productivity</td>
<td>Profit</td>
<td>Previous Productivity</td>
<td>Profit</td>
</tr>
<tr>
<td>Current Productivity</td>
<td>.65**</td>
<td>.41**</td>
<td>.48**</td>
<td>.37**</td>
</tr>
<tr>
<td>Current Profit</td>
<td>.40**</td>
<td>.59**</td>
<td>.38**</td>
<td>.41**</td>
</tr>
<tr>
<td>Previous Productivity</td>
<td>.47**</td>
<td>.38**</td>
<td>.59**</td>
<td>.37**</td>
</tr>
<tr>
<td>Previous Profit</td>
<td>.37**</td>
<td>.42**</td>
<td>.38**</td>
<td>.52**</td>
</tr>
</tbody>
</table>

Note: Figures in bold are for directly corresponding subjective and objective performance measures.

*p < .01  **p < .001
whereas for multi-site companies the measure of the practices is for one (typical) site but that for performance is for the company as a whole. Differentiating the sample in this way enables further assessment of the consistency of the findings.

The findings bearing on convergent and discriminant validity are presented in Table 3. For both single- and multi-site companies the relationships between the directly corresponding subjective and objective performance measures (see on-diagonal coefficients in bold) are the strongest. For instance, the correlation between subjective and objective current productivity for single-site companies is .65 ($p < .001$) and that for multi-site companies is .52 ($p < .001$). Inspection suggests that the coefficients for the single-site companies (from .52 to .65) are higher than those for multi-site companies (.38 to .52), but statistical tests show that in no instance do directly corresponding coefficients differ statistically significantly. We conducted those tests based on the difference between the respective $z$ coefficients, even though we used a nonparametric correlation measure. This is justified because the Spearman’s rank-order correlation “is actually the Pearson $r$ that would be obtained from correlating the rank values for the same basic data... the sampling distribution may be assumed to be normal and we may estimate a $z$ ratio” (Guilford & Fruchter, 1978, p. 295).

The evidence of convergent validity is complemented by that for discriminant validity. This is shown by the smaller (off-diagonal) correlations for the relationships between nonequivalent measures, be those between alternative periods (current or previous) or aspects (productivity or profit) of performance, or both. Positive correlations among these nonequivalent measures are of course to be expected, because, in practice, productivity contributes to profit and performance tends to be consistent over time. Nonetheless, in neither subsample does any one coefficient for the relationship between different constructs measured by the same method (i.e., subjective or objective) exceed even the lowest coefficient for the relationship between subjective and objective measures of the same construct. We formally assessed for convergent coefficients being greater than the off-diagonal (discriminant) ones as in Study 1 (using Hotelling’s $t$-test). These tests (for each of the 4 convergent coefficients against each of the 6 nonconvergent ones involving one or other of the variables in the convergent correlations) showed statistically significant ($p < .05$) effects in 16 of the 24 cases in the single-site subsample, and in 20 of the 24 cases in the larger multi-site subsample. Thus, discriminant validity is evident not only in the overall pattern of the results but at the level of individual coefficients.
<table>
<thead>
<tr>
<th>Performance</th>
<th>Single-site companies (n = 76–83)</th>
<th>Multi-site companies (n = 264–278)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Productivity</td>
<td>Profit</td>
</tr>
<tr>
<td>Just-in-time</td>
<td>.28**</td>
<td>.31**</td>
</tr>
<tr>
<td>Total quality management</td>
<td>.27*</td>
<td>.30**</td>
</tr>
<tr>
<td>Empowerment (job enrichment)</td>
<td>.24*</td>
<td>.27*</td>
</tr>
<tr>
<td>Learning culture (skills enhancement)</td>
<td>.36**</td>
<td>.40**</td>
</tr>
<tr>
<td>Supply-chain partnering</td>
<td>.15</td>
<td>.15</td>
</tr>
<tr>
<td>Business process reengineering</td>
<td>.12</td>
<td>.13</td>
</tr>
<tr>
<td>Team-based work</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>Outsourcing</td>
<td>.19</td>
<td>.22</td>
</tr>
<tr>
<td>Total productive maintenance&lt;sup&gt;1&lt;/sup&gt;</td>
<td>.41**</td>
<td>.37**</td>
</tr>
</tbody>
</table>

<sup>1</sup> Applicable only to the subsample of manufacturing companies (n = 242)

*<i>p < .01</i>  **<i>p < .001</i>
The question of construct validity is again addressed by assessing whether subjective and objective measures would lead to different conclusions concerning the relationship between management practices and performance. As for the previous study, the overall pattern of the findings is unambiguous (see Table 4). Whether focusing on productivity or profit, in single- or multi-site companies, the rank-order correlation coefficients for the two types of measure are remarkably similar. The differences between the coefficients are typically of the order of .00 to .04, and in no instance are they greater than .06. We formally assessed for this equivalence in the same way as described above, using Hotelling's t-test. No difference even approaching statistical significance was found.

Considering the findings using conventional statistical significance levels would lead to an equally strong conclusion. In the 23 instances in which the subjective performance measure shows no statistically significant relationship with a practice, neither does the corresponding objective performance measure. Correspondingly, in the 12 of the 13 cases where the subjective measure does correlate statistically significantly with a practice, so too does the objective one. The single exception is for the relationship of just-in-time with productivity in multi-site companies, where the subjective productivity is statistically significantly related to just-in-time (.25, \( p < .01 \)) but the objective measure is not (.23, ns). However, this is no more than a marginal difference around the arbitrary probability threshold adopted. Moreover, where the alternative subjective performance measures (i.e., productivity and profit) correlate differentially with a practice (e.g., outsourcing and total productive maintenance), so do the corresponding objective performance measures.

Discussion

This second study is more comprehensive than Study 1 in that it has two subsamples, measures two different aspects of performance, and covers more management practices. The opportunity to register discordant results is consequently greater. Yet, the findings consistently demonstrate convergent, discriminant, and construct validity. Once again, the evidence for convergent validity is arguably the least convincing, as subjective measures do not correlate as strongly as might be expected with their objective counterparts.

Nonetheless, the evidence for construct validity is highly consistent. Even though data on the practices and subjective performance were obtained from the same source for all but 61 of the companies, there is no evidence of common-method variance. The relationships of the use of the management practices with subjective performance are no greater than those with objective performance. Further analyses excluding cases
where subjective performance was provided by a second nominated respondent made no material difference to the results. (These results are available from the second named author, upon request.)

General Discussion

In this paper, we combined findings from two studies, and three samples, to examine the validity of self-report measures of company financial performance. In all cases, there was evidence of convergent validity, as the subjective performance measures were shown to be related to their objective counterparts. Equally there was consistent evidence of discriminant validity, in that the relationships between corresponding subjective and objective measures were systematically stronger than the relationship between different performance constructs measured using the same method (either subjective or objective).

What makes these results especially significant, however, is the evidence for construct validity. Findings relating the use of management practices to subjective measures of performance were essentially equivalent to those for objective performance. Across the studies, we examined 44 relationships between various management practices and both subjective and objective performance. In no case was the correlation between a given practice and a measure of subjective performance statistically significantly different from that for the relationship of that same practice with the corresponding objective measure of performance. Looking at this another way, through the lens of conventional statistical significance testing, in each and every instance of the 13 cases where a practice was statistically significantly related to objective performance at $p < .01$, its relationship with subjective performance was also statistically significant. Conversely, in the 31 cases where a nonsignificant relationship between a practice and objective performance was found, 29 of the corresponding relationships involving subjective performance were also nonsignificant. Overall, judged against their objective counterparts, subjective performance measures showed a 95% success rate.

The two discrepancies, however, were "false positives," in that the subjective measure showed a statistically significant relationship when the corresponding objective measure did not. This might indicate a bias in subjective performance measures that overestimates effects, consistent with the common-method variance argument. Three considerations weaken that interpretation. First, one of the false positives was in Study 1 where, because the measures of the practices and subjective performance came from separate sources, common-method variance was not a risk. Second, the difference between the coefficients in both studies, and especially in Study 2 where the risk of common-method variance
bias applies, was slight and statistically insignificant, with the discrepancy arising because the coefficient for subjective performance was just below the arbitrary significance level and the objective one just above. Third, there is no tendency in the results more generally for coefficients involving the subjective measures to be greater than those for the objective ones. To confirm this interpretation, we conducted a Wilcoxon signed-rank test using the correlation coefficients as data points. This yielded a z value of 1.35 (ns), showing no statistically significant difference between the coefficients for corresponding subjective and objective performance measures.

The question of convergent validity, however, warrants further consideration because the direct correlations between the subjective and objective performance measures were not especially strong. Across the studies they were typically of the order of .40 to .60, which leaves considerable independent variation in each. In part this may arise because the construct in the subjective measure was not identical to that operationalized in the objective one. This was certainly the case in Study 1, where the subjective measure concerned performance in general, whereas the objective one focused on profit. Consequently, in rating performance, respondents could reasonably have taken into consideration factors other than the specific financial one incorporated into the objective measure, reflecting the broader view of company performance that Rogers and Wright (1998) recommend. The argument of nonequivalence, however, holds less weight for Study 2, where the subjective and objective measures of productivity and profit were more closely aligned. Correspondingly, the evidence for convergent validity was stronger in Study 2. This pattern of stronger convergent validity as a function of alignment between the content of subjective and objective measures is discernible in the previous literature. Using a general subjective performance measure, Baer and Frese (2003) recorded a correlation of .41 with an objective index of return on assets; whereas Guthrie (2001) specifically asked respondents to report sales per employee and found a correlation of .81 with the corresponding index from audited accounts.

Another reason for the relatively modest convergent validity coefficients observed, however, may lie in error within the objective measures. In our discussion so far we have emphasized potential sources of invalidity in subjective measures and taken the objective measure as the criterion, as if this were unequivocally valid as a measure of company performance. Yet, that assumption should be challenged. Aside from the recent well publicized instances of fraudulent financial recording of company performance, which are no doubt rare, there are many perfectly reasonable and accepted reasons for audited accounts not to fully reflect actual performance. It is common practice, for example, to bring
forward planned expenditure in good years to offset tax or for cash flow reasons, and, conversely, to delay investment in bad years, thus carrying over cost in the light of performance. Similarly, companies can invest in their future, say through running an executive development program or buying new equipment, with the cost charged in one particular year but the benefits spread over several years. More generally, as Smith (1996) has powerfully illustrated, a wide range of profit levels can be reported in company accounts depending upon the assumptions made on such matters as depreciation rates and the cost of stock options. Thus, although it is clear from our findings that subjective and objective measures of financial performance have much in common, we cannot assume that either is error free. The consequence of such error in both measures, of course, is that our findings for convergent validity will be attenuated, and the strength of the true underlying relationship is likely to be greater than that actually recorded.

It is interesting that the findings in this study closely parallel those for subjective and objective measures of individual employee performance. Bommer et al.'s (1995) meta-analysis of 50 studies also found only moderate evidence for convergent validity in that the overall corrected mean correlation between the two types of measure was .39 (.32 observed). At the same time, however, they summarize previous evidence of construct validity (e.g., McEvoy & Cascio, 1989; Nathan & Alexander, 1988; Ones, Viswesveran, & Schmidt, 1993), which corresponds to our findings for company performance. Thus, Bommer et al. (1998) note: "researchers have explicitly looked for differences in relationship strength between multiple independent variables with subjective and objective measures. But contrary to expectation, in practically no cases have significant relationship differences been found" (p. 588). Our findings with regard to convergent validity are arguably stronger, especially in Study 2 where the subjective and objective measures of company performance are better aligned conceptually. Given that reporting on company performance is likely to be less threatening than reporting on individual employee performance (e.g., because of fewer personal pay or employment security implications), and because there are public records of key aspects of company performance to both inform reports and against which they could be checked, we would expect subjective measures of company performance to have greater convergent validity than their employee performance counterparts.

Bommer et al.'s (1998) findings concerning the equivalence of findings for subjective and objective measures of employee performance also have a parallel within the empirical literature on human resource management (HRM) and organizational performance. Those organizations characterized by more "high performance" or "human capital enhanc-
ing" HRM practices have been found to have better performance, irrespective of the use of objective (e.g., Huselid, 1995) or subjective (e.g., Youndt et al., 1996) performance measures. Equally, the absence of such a relationship has been reported for both objective (Guest et al., 2003) and subjective (Wood & de Menezes, 1998) performance measures.

The main limitation of the present studies is that they focus on accountancy-based performance measures of productivity and profit. In the literature there is a much wider range of company measures considered, with market-based measures of shareholder value also being especially pertinent to human resource management approaches linked to resource-based theories of the firm. It remains to be seen whether our findings for productivity and profit generalize to such other measures.

There are three main implications of this research. First, the degree of equivalence between the findings for subjective and objective measures means we can have some confidence in findings from studies so far that have been based on subjective company performance. Correspondingly, the findings support the use of subjective performance measures in future studies where objective ones may not be feasible, for instance for certain sectors of the economy (e.g., public sector or charitable organizations) or for certain levels of analysis that do not have separate objective measures (e.g., plants, business units, or divisions within companies). There is little evidence from the current findings to suggest that such studies would lead to erroneous conclusions, though further replication of our results is desirable.

The second, and most important, implication is that there is a need to improve both types of measure. Doubtless there is room to improve objective measures, by investigating the accounting decisions made for any given reporting period and making any adjustments so indicated. In large-scale studies, however, that may not be a realistic proposition because the detailed historical financial information required is often not readily available. With regard to subjective measures, there is an obvious opportunity to systematically examine the relative merits of different approaches. Given our findings, and those in the literature to which we have referred, it is worthwhile examining the extent to which subjective measures could be improved by asking specific questions that directly correspond to the objective performance index of interest, investigating the relative validity of subjective measures of different kinds (e.g., accountancy as opposed to market value ones), providing respondents with prior warning of the areas to be covered, and focusing on particular categories of respondent (perhaps finance managers or CEOs rather than production or HR managers). Moreover, evidence needs to be obtained about the relative merits of absolute compared with relative judgments. These suggestions point to areas for future research.
The third and final implication is that, wherever possible, investigators should use both subjective and objective measures of performance within studies. Given each type of measure will contain its own error, more reliable estimates of performance may be obtained by combining them. At the same time, equivalent results from both types of measure would add weight to any substantive findings, and differential findings alert investigators to underlying problems.

REFERENCES


