Overall Job Satisfaction: How Good Are Single-Item Measures?

John P. Wanous and Arnon E. Reichers
Fisher College of Business, Ohio State University

Michael J. Hudy
Ohio State University

A meta-analysis of single-item measures of overall job satisfaction (28 correlations from 17 studies with 7,682 people) found an average uncorrected correlation of .63 (SD = .09) with scale measures of overall job satisfaction. The overall mean correlation (corrected only for reliability) is .67 (SD = .08), and it is moderated by the type of measurement scale used. The mean corrected correlation for the best group of scale measures (8 correlations, 1,735 people) is .72 (SD = .05). The correction for attenuation formula was used to estimate the minimum level of reliability for a single-item measure. These estimates range from .45 to .69, depending on the assumptions made.

In general, single-item measures can be divided into two categories: (a) those measuring self-reported facts, such as years of education, age, number of previous jobs, and so on, and (b) those measuring psychological constructs, such as job satisfaction. Measuring the former with a single item is commonly accepted practice. However, the use of single-item measures for psychological constructs is typically discouraged, primarily because they are presumed to have unacceptably low reliability. In fact, the use of single-item measures in academic research is often considered a “fatal error” in the review process.

There are exceptions to the norm of using only scales to measure psychological constructs, however. If the construct being measured is sufficiently narrow or is unambiguous to the respondent, a single-item measure may suffice, as pointed out by Sackett and Larson (1990). For example, most expectancy theory researchers use a single item to measure the perceived probability that effort leads to performance (e.g., Ilgen, Nebeker, & Pritchard, 1981).

For more complex psychological constructs (e.g., personality), it is usually recommended that scales with multiple items be used. However, there is a middle ground between the extremes of simple (expectancy) and complex (personality) constructs where measurement with a single item may also be acceptable. Overall job satisfaction appears to be an example, particularly because it has a long history of single-item measures, beginning with the Faces Scale over 40 years ago (Kunin, 1955).

Scarpello and Campbell (1983) concluded that a single-item measure of overall job satisfaction was preferable to a scale that is based on a sum of specific job facet satisfactions. In the years since that article was published, however, the legitimacy of single-item measures has not increased. There are probably several reasons for this. First, it is frequently said that one cannot estimate the internal consistency reliability of single-item measures, and this alone is sometimes believed to be a sufficient reason to limit or avoid their use. A second reason may be the increased use of structural equation models that require estimation of the reliable variance.

A third reason is that readers of Scarpello and Campbell (1983) may not have been convinced that the moderate correlations they reported between the single-item and a scale of job facet satisfactions were due to construct definitions, as they argued. Rather, readers may have concluded that low reliability for the single-item measures was a better explanation.

The Present Study

The present study is a meta-analysis (Hunter & Schmidt, 1990) of research in which single-item measures of overall job satisfaction are correlated with scales measuring overall job satisfaction. This was done for two reasons. First, the average correlation is more representative of the population correlation than a correlation that is based on a single sample. Second, two pieces of information from the meta-analysis are used to estimate the
reliability of single-item measures of overall job satisfaction by using a well-known formula (i.e., the correction for attenuation due to unreliability) in a new way. Specifically, the average observed correlation between the single-item and the scale measures of job satisfaction can be put into this formula, along with the estimated average reliability of the scales used to measure overall job satisfaction. By adding a value for the assumed underlying construct correlation between these two measures of overall job satisfaction, it is possible to solve the equation for its missing value, that is, the minimum reliability of the single-item measure. This procedure is described below. (See also Wanous & Reichers, 1996b.)

The well-known formula for the correction for attenuation can be found in most textbooks on psychometrics (e.g., Nunnally & Bernstein, 1994, p. 257):

$$\hat{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} \cdot r_{yy}}}$$

where $r_{xy} =$ correlation between variables $x$ and $y$, $r_{xx} =$ reliability of variable $x$, $r_{yy} =$ reliability of variable $y$, and $\hat{r}_{xy} =$ the assumed true underlying correlation between $x$ and $y$ if both were measured perfectly.

This formula is most commonly applied to situations where variables $x$ and $y$ are different constructs, for example, job satisfaction and job performance. However, the formula can also be applied when both measured variables represent the same construct (i.e., job satisfaction, as done here). When this occurs, Nunnally (1978) reported the following:

The correlation between two such tests would be expected to equal the product of the terms in the denominator and consequently $\hat{r}_{xy}$ would equal 1.00. . . . If $\hat{r}_{xy}$ were 1.00, $r_{xy}$ would be limited only by the reliabilities of the two tests: $r_{xy} = \sqrt{r_{xx} \cdot r_{yy}}$. (p. 220)

Method

Literature Search

The following search procedures were used. First, a computerized search of the American Psychological Association's PsycLit database was conducted. Second, a manual search of studies cited in various published empirical articles and an unpublished master's thesis (Thomas, 1993) on this topic was conducted. Third, a second computerized search was done with the Business Abstracts database.

Criteria for Inclusion

Studies selected for this meta-analysis reported correlations between two measures of overall job satisfaction, one of which was a single-item measure. The second measure could be either a sum of job facet satisfactions or the sum of several items focused on overall job satisfaction. This led to the exclusion of two published studies (Ferratt, 1981; Kalleberg, 1974) because their multiple-item measures were judged to be poor measures of overall job satisfaction. This is because an indirect measure of job satisfaction was used ("should be—is now"), which involves calculating a raw discrepancy score. This measure of job satisfaction has been severely criticized (Wall & Payne, 1973). Although judgment calls about the inclusion or exclusion of data can affect a meta-analysis (Wanous, Sullivan, & Mallnak, 1989), omitting these two studies had no effect here, on the basis of a preliminary analysis.

Recorded Information

The sample size, type of job satisfaction measure, reliability of the scale measure, and the correlation between the single item and the scale were recorded for each study. The type of single-item measure was subdivided into (a) 15 correlations that used the Faces Scale, and (b) 13 correlations that used some other type of single-item measure. The scales used to measure overall job satisfaction were subdivided into (a) 8 correlations that used a sum of items with a focus on overall job satisfaction, (b) 9 correlations that used a sum of facets concerning only the work itself, and (c) 11 correlations that used a sum of job facet satisfactions.

Meta-Analysis Calculations

A software package developed by Switzer (1991) was used to make the meta-analytic calculations. This program uses the Hunter and Schmidt (1990) interactive procedure to produce both observed and corrected means and standard deviations for a set of correlations. The program also uses a bootstrap resampling method to re-estimate the parameter estimates of the mean and variance for comparison to the Hunter and Schmidt (1990) estimates. Furthermore, the bootstrap method yields 95% confidence intervals for each of these estimates (see Switzer, Paese, & Drasgow, 1992, for a discussion of the bootstrap method).

The specific corrections made here include (a) sampling error, and (b) criterion reliability, in this case the reliability of the scales, which averaged .88 for the six studies reporting this information.

Results

The Appendix shows the results of the literature search. Studies are listed alphabetically. In all, there were 17 different samples from 16 sources that reported a total of 28 correlations. There are two ways to report the sample size: (a) the number of distinct persons in the 17 samples ($N = 4,568$; double counting of people is eliminated because some studies report multiple correlations), and (b) the number of persons for all 28 correlations ($N = 7,682$, which includes double counting).

Table 1 shows the results of the meta-analysis. The mean observed correlation is .63 ($SD = .09$). When this is corrected for unreliability, the correlation is .67 ($SD = .08$). Total artifactual between-studies variance was 24%
of the variance, of which 18% was attributed to sampling error.

As a check on the possible effect of double-counting people when multiple correlations are reported, the calculations were recalculated on the 17 distinct samples. For those studies reporting more than one correlation, the results were averaged and weighted by the average sample size. The results are quite similar to the results using all 28 correlations. The corrected mean correlation is .66 (SD = .08) for the 17 different samples.

Because 76% of the between-studies variance could not be accounted for by measurable artifacts, a search for moderators was undertaken. Two moderators were considered: (a) how the single item was measured: Faces vs. another type of single-item scale and (b) how the scales were measured: a sum of facet satisfaction, a sum of items focused on overall satisfaction, and a sum of facets concerning only the work itself, primarily the Work facet of the JDJ.

The type of single-item measure does not moderate the overall correlation. However, the type of scale used to measure overall job satisfaction does. Specifically, the difference between using a scale focused only on the work itself versus a scale that is based on items focused on overall satisfaction has the classical characteristics of a moderator. That is, there is a mean difference between these two groups (r = .63 vs. .72) and the standard deviations for each group are smaller than when they are combined. The third scale measure, sums of facet satisfactions, had properties similar to the overall results.

Table 1 also shows the bootstrap re-estimates of the mean and standard deviation from the meta-analysis and the 95% confidence intervals for each of these re-estimates. This procedure was developed (Switzer, Pease, & Drasgow, 1992) to deal with the uncertainty associated with such estimates, particularly when the number of correlations is relatively small. However, in the present study, the two methods yielded almost identical results.

Finally, it is possible to estimate the minimum reliability of a single-item measure using the correction for attenuation formula, as described above. Several different estimates were made, but in all cases the average observed correlations were used, not the corrected ones, because that would have resulted in double-correction and inflated estimates.

A total of eight estimates of the minimum reliability of a single-item measure are shown in Table 2. These are based on (a) which observed mean correlation was used, (b) which assumption about the underlying construct correlation was used, and (c) which scale reliability was used. First, the mean observed correlation based on all 28 that were found in our review (.63) was used, as is the mean correlation for the subgroup using the best scale measure of overall job satisfaction, that is, those scales using items that have a focus on overall satisfaction (.67).

<table>
<thead>
<tr>
<th>Assumptions about the construct correlation and about scale reliability</th>
<th>Average observed correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies (r = .63)</td>
<td>Best scale (r = .67)</td>
</tr>
<tr>
<td>Reliability = .88 (observed)</td>
<td>.45</td>
</tr>
<tr>
<td>Reliability = .80</td>
<td>.50</td>
</tr>
<tr>
<td>Assume true r = .90</td>
<td>Reliability = .88 (observed)</td>
</tr>
<tr>
<td>Reliability = .80</td>
<td>.61</td>
</tr>
<tr>
<td>$M$</td>
<td>.63</td>
</tr>
</tbody>
</table>
challenged this assumption. As a result, estimates based on a construct correlation of .90 are also shown. Third, only 6 of 17 studies reported scale reliability information, and it averaged .88, which is fairly high. Because this average may not be representative, an alternative average reliability of .80 was also used.

The estimates shown in Table 2 range from a low of .45 to a high of .69, with an average of .57. Remember, these estimates are of the minimum reliability for a single item. The actual reliability could be higher, but it cannot be lower nor can it be estimated with any greater precision than a minimum value. The effect of each of the three assumptions on the estimated minimum values approximately equal, as shown in Table 2.

Discussion

The observed correlations between single items and scales averaged .63, and the corrected mean correlation of .67 further indicated convergent validity. This average is much higher than the results reported by Scarpello and Campbell (1983) and seems to bolster their argument that a single-item measure of overall job satisfaction is acceptable.

The average correlation is highest for the scales composed of items with an overall job satisfaction focus, as compared with a sum of facets or the Work scale from the JDI. This is what one would expect to see because the focus of the items is directly on overall job satisfaction, and thus most similar to the wording of the single-item measures.

The minimum level of reliability for single-item job satisfaction measures was also estimated by using the standard correction for attenuation formula in a somewhat novel way. Reasonable estimates of this minimum value are those based on assuming that the true underlying construct correlation is .90 rather than 1.00, combined with using the best measures available. These estimated minimum reliabilities are .63 and .69, depending on the assumption about the reliability of the scales. In fact, assuming the reliability is closer to .80 than .88 is also realistic because an average reliability of .82 has been reported for some job satisfaction scales (Fried & Ferris, 1987). Thus, a minimum estimated reliability for the single-item measure close to .70 is reasonable.

It is also interesting to note that differences among single-item measures had no effect on the meta-analysis results. However, differences in the ways that scales were measured did make a difference in the results. Because of this difference, it seems reasonable to conclude that the single-item measures are more robust than the scale measures of overall job satisfaction.

Despite the above data and arguments, there are still good reasons for preferring scales to single items. Nothing reported thus far should be interpreted as questioning the use of well-constructed scales in comparison to a single-item measure. It should be interpreted, however, as a case for the acceptability of single-item measures when either the research question implies their use or when situational constraints limit or prevent the use of scales.

One example of a research question suggesting the use of a single-item measure is the measurement of change in overall job satisfaction. Suppose that a single question is asked of respondents: "How much did your overall job satisfaction change in the last year?" Such a single-item measure might be preferable to a scale, as the following example illustrates. Imagine a situation where Person A has an average scale score of 3 by answering 3 to each of 5 questions. Person B also has an average scale score of 3 but answered two questions with 1, two more with 5, and the last with 3. These differences are ignored if a scale score is used.

There may also be practical limitations favoring the use of a single-item measure. For example, if space on a questionnaire is limited, researchers have to make choices about how many items will be included. Specifically, the Minnesota Satisfaction Questionnaire (Short Form) has 20 job-facet items. If this popular scale is used only to measure overall job satisfaction, 19 extra questions must be answered.

There may be shorter alternatives to using a long list of job facets, such as scales that are focused on overall job satisfaction (Brayfield & Rothe, 1951; Hoppock, 1935; Taylor & Bowers, 1972). However, these are all quite wordy and consume precious space on a survey. The shortest of these is probably the five-item General Satisfaction Scale from the Job Diagnostic Survey (JDS; Hackman & Oldham, 1975).

There might also be relevant cost considerations that would lead researchers to prefer a single-item measure. For example, our university telephone poll charges $60 per item per sample (faculty, staff, students). Therefore, using the shortest of the scales (five items from the JDS) would add $240 to the cost of measuring overall job satisfaction for one sample and $720 for all three.

Finally, there may be issues of face validity. In particular, respondents may resent being asked questions that appear to be repetitious. This problem would be worse in an organization with poor employee relations. From a management perspective, a single item is usually easier to understand than a scale score, which might appear as academic nit-picking.

In summary, if neither the research question nor the research situation suggest the use of a single-item job satisfaction measure, then choosing a well-constructed scale makes sense. However, if the use of a single item is indicated, researchers may do so in the knowledge that they can be acceptable. The use of single-item measures
should not be considered fatal flaws in the review process. Rather, their appropriateness for a particular piece of research should be evaluated. Future research should be directed toward examining the validity and reliability of other single-item measures.

References

References marked with an asterisk indicate studies included in the meta-analysis.


(Appendix follows on next page)
### Studies Used in the Meta-Analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>α</th>
<th>Single item</th>
<th>Scale</th>
<th>t²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adler &amp; Golan (1981)</td>
<td>82</td>
<td>.98</td>
<td>Faces</td>
<td>Sum of facets</td>
<td>.74 (JDI)</td>
</tr>
<tr>
<td>Brief &amp; Roberson (1989)</td>
<td>144</td>
<td>.92</td>
<td>Faces</td>
<td>Sum of facets</td>
<td>.71 (JDI)</td>
</tr>
<tr>
<td></td>
<td>144</td>
<td>.89</td>
<td></td>
<td></td>
<td>.71 (MSQ-SF)</td>
</tr>
<tr>
<td>Dunham &amp; Herman (1975)</td>
<td>113</td>
<td></td>
<td>Faces</td>
<td>Overall</td>
<td>.75 (Hoppock Job Satisfaction Blank)</td>
</tr>
<tr>
<td>Horn et al. (1979)</td>
<td>373</td>
<td></td>
<td>Faces</td>
<td>Work facet</td>
<td>.53 (JDI)</td>
</tr>
<tr>
<td>Ironson et al. (1989)</td>
<td>227</td>
<td></td>
<td>Item (A)</td>
<td>Overall</td>
<td>.75 (JIG) .65 (B-R) .68 (JDI)</td>
</tr>
<tr>
<td></td>
<td>227</td>
<td></td>
<td>Item</td>
<td>Work facet</td>
<td>.67 (JIG) .60 (B-R) .59 (JDI)</td>
</tr>
<tr>
<td>LaRocco (1983)</td>
<td>260</td>
<td>.92</td>
<td>Item</td>
<td>Overall</td>
<td>.58 (JDS)</td>
</tr>
<tr>
<td>Miller et al. (1979)</td>
<td>235</td>
<td>.88</td>
<td>Faces</td>
<td>Work facet</td>
<td>.63 (JDI)</td>
</tr>
<tr>
<td></td>
<td>225</td>
<td></td>
<td>Faces</td>
<td>Work facet</td>
<td>.59 (JDI)</td>
</tr>
<tr>
<td>Mossholder et al. (1988)</td>
<td>220</td>
<td>.89</td>
<td>Item</td>
<td>Sum of facets</td>
<td>.71 (MSQ-SF)</td>
</tr>
<tr>
<td>O'Reilly &amp; Roberts (1973)</td>
<td>139</td>
<td></td>
<td>Faces</td>
<td>Work facet</td>
<td>.45 (JDI)</td>
</tr>
<tr>
<td>O'Reilly &amp; Roberts (1978)</td>
<td>562</td>
<td></td>
<td>Faces</td>
<td>Work facet</td>
<td>.59 (JDI)</td>
</tr>
<tr>
<td>Rosse &amp; Hulin (1985)</td>
<td>42</td>
<td>.81</td>
<td>Faces</td>
<td>Work facet</td>
<td>.44 (JDI)</td>
</tr>
<tr>
<td>Russell &amp; Farrar (1978)</td>
<td>507</td>
<td>.91</td>
<td>Faces</td>
<td>Sum of facets</td>
<td>.63 (JDI)</td>
</tr>
<tr>
<td>Scarpello &amp; Campbell (1983)</td>
<td>185</td>
<td></td>
<td>Item</td>
<td>Sum of facets</td>
<td>.32 (MSQ)</td>
</tr>
<tr>
<td></td>
<td>185</td>
<td></td>
<td>Item (Yes-No)</td>
<td></td>
<td>.42 (MSQ)</td>
</tr>
<tr>
<td>Smith et al. (1969)</td>
<td>80</td>
<td></td>
<td>Faces</td>
<td></td>
<td>.54 (JDI)</td>
</tr>
<tr>
<td>Wansou &amp; Lawler (1972)</td>
<td>208</td>
<td></td>
<td></td>
<td>Sum of facets</td>
<td>.60</td>
</tr>
<tr>
<td>Wansou &amp; Reichers (1996a)</td>
<td>969</td>
<td>.86</td>
<td>Item</td>
<td>Sum of facets</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>966</td>
<td>.77</td>
<td></td>
<td></td>
<td>.68</td>
</tr>
</tbody>
</table>

*Single item measures include (a) the Faces scale—Faces, (b) a numerical item—Item, (c) an adjective item—Item (A), and (d) an item with a yes–no response—Item (Yes–No). Scale measures include (a) a sum of facet satisfaction items—Sum of facets, (b) a sum of overall-focused items—Overall, and (c) a sum of facet satisfaction items concerning only the work itself—Work facet. JDI = Job Descriptive Index; MSQ = Minnesota Satisfaction Questionnaire; MSQ-SF = Short Form of the MSQ; B-R = Brayfield-Rothe Scale.

**Received May 23, 1996**

**Revision received September 13, 1996**

**Accepted September 17, 1996**