DEVELOPMENT OF A COMPACT MEASURE OF JOB SATISFACTION: THE ABRIDGED JOB DESCRIPTIVE INDEX

JEFFREY M. STANTON, EVAN F. SINAR, WILLIAM K. BALZER,
AMANDA L. JULIAN, PAUL THORESEN, SHAHNAZ AZIZ,
GWENITH G. FISHER, AND PATRICIA C. SMITH
Bowling Green State University

The Job Descriptive Index is a popular measure of job satisfaction with five subscales containing 72 items. A national sample (n = 1,534) and a sample of university workers (n = 636) supported development of an abridged version of the Job Descriptive Index (AJDI) containing a total of 25 items. A systematic scale-reduction technique was employed with the first sample to decide which items to retain in each scale. The abridged subscales were then tested in the second sample. Results indicated that the relationships among the five abridged subscales and between the five abridged subscales and other measures were substantially preserved.

Job satisfaction has a long and fruitful history in the organizational sciences. Balzer et al. (1997) suggested that humanitarian, economic, and theoretical concerns have all contributed to the continuing importance and popularity of job satisfaction. Humanitarian concerns spring from the ethical imperative felt by many managers to provide an acceptable, agreeable, and pleasant work environment for employees. The economic perspective is exemplified in recent research by Organ (1988; Organ & Ryan, 1995) and by Ostroff (1992) that has helped to clarify the relation between job satisfaction and organizational performance. From a theoretical perspective, models of work motivation treat job satisfaction as a critical job attitude that can help predict behavior (Steers, Porter, & Bigley, 1996). Together, these orientations have provided the impetus for both applied and academic research into job satisfaction.

The first contemporary measure of job satisfaction, published by Hoppock in 1935, was a 4-item measure of general job satisfaction. Dozens

Educational and Psychological Measurement, Vol. 61 No. 6, December 2001 1104-1122

of measures followed to assess both general job satisfaction and specific facets of satisfaction (Cook, Hepworth, Wall, & Warr, 1981). Among these, the Job Descriptive Index (JDI) (Smith, Kendall, & Hulin, 1969) provides five subscales that measure different facets of job satisfaction. Each JDI facet scale contains either 9 or 18 adjectives or short adjectival phrases describing various aspects of the respondent's work experiences—the work itself, pay, opportunities for promotion, supervision, and coworkers. Although the instrument's brief response format (yes, no, or ?), item brevity, and low required reading level simplify the respondent's task, the instrument's 72 items take a substantial amount of space on a survey instrument and require several minutes to complete.

Two trends in organizational research have exacerbated the trouble with lengthy self-report measures such as the JDI. First, it is clear from examples of recent research on job satisfaction (e.g., Ganzach, 1998) that job satisfaction is rarely measured in isolation but is instead measured alongside numerous other constructs. Adequately measuring multiple constructs virtually guarantees a lengthy survey instrument. In addition, Rogelberg (1998; Rogelberg & Luong, 1998) has documented the finding that many organizational members feel that they are "oversurveyed" and that these feelings increase the likelihood of nonresponse. The multivariate nature of modern organizational research and the apparent survey fatigue of organizational members make a brief but psychometrically sound measure of job satisfaction desirable. The present article reports two validation studies aimed at producing a brief measure of each of the JDI's five facets of job satisfaction: the Abridged Job Descriptive Index (AJDI).

A Case for Continued Development of the JDI

The JDI has been described as the most popular and widely used measure of job satisfaction (Buckley, Carraher, & Cote, 1989; DeMeuse, 1985; Zedeck, 1987). The instrument has been translated into nine different languages and administered in at least 17 countries. Development of the JDI began in the early 1960s and was first marked in the literature by the publication of Locke, Smith, Kendall, Hulin, and Miller (1964). That article presented initial findings from a validation effort that eventually included 21 different industrial samples in 18 organizations. These organizations included a variety of industries, and the respondents within them covered the organizational hierarchy from top management to line workers (Smith et al., 1969). A more detailed history of these activities appeared in Smith and Stanton (1999).

In later work on the JDI, Smith, Smith, and Rollo (1974) confirmed the factor structure of the JDI in a racially diverse sample. Johnson, Smith, and

Tucker (1982) showed that the JDI response format was not inferior to a Likert-scaled format. Ironson, Smith, Brannick, Gibson, and Paul (1989) developed the initial version of the multiitem Job in General (JIG) scale, which added a psychometrically sound, global satisfaction measure to the facet measures. Additional research on the JDI and related measures resulted in the publication of a second book, simply titled *Job Satisfaction* (Cranny, Smith, & Stone, 1992). Researchers have updated the item content, validity evidence, and national norms in the three decades since JDI's original publication (Smith et al., 1987). The most recent update of JDI item content and a complete renorming of the instrument occurred in 1997 (Balzer et al., 1997). A computerized literature search of journal articles, book chapters, and dissertations in psychology indicated that the various revisions of the instrument have been used in more than 300 published and unpublished research projects to date. The instrument's continuing popularity with organizational researchers, in conjunction with the pressing need for shorter measures in organizational research, warranted the development of an abridged version of the instrument.

Overview of the Scale-Reduction Strategy

Our overriding goal in reducing the length of the JDI was to preserve the qualities that have made the instrument useful to organizational researchers. For example, researchers have ascertained that the instrument has yielded scores with high internal consistency that are usable for a variety of respondent populations and predictive of a number of organizationally relevant outcomes such as intentions to quit the organization (Balzer et al., 1997). We clustered the qualities of the JDI into three domains: score validity, psychometric qualities, and user features.

In regard to validity evidence, we intended to preserve as much as possible the magnitudes of the correlations between JDI facet scores and external criteria (e.g., general job satisfaction, intention to quit). For psychometric qualities, we wanted to maintain an acceptable level of internal consistency for scores on each scale, ensure inclusion of items that covered different degrees of job satisfaction, and try to avoid exacerbating the skewness exhibited by many distributions of job satisfaction. For user features, we wanted to maintain the "face validity" of the instrument, include both positively worded and negatively worded items, avoid redundancy of item content, and most important, decrease the administration time. Note that these latter issues are somewhat subjective and require a degree of professional judgment to fulfill, especially insofar as they conflict with the psychometric and validity goals. For example, quicker administration time requires fewer items, whereas attaining higher internal consistency typically demands more items.

Because some of these goals were interrelated and a few of them were potentially contradictory, we needed to systematically mediate trade-offs

between the various issues. To this end, we developed a methodology to help quantify a broad set of item qualities. For example, because an item with a large corrected item-total correlation typically increases the internal consistency of scale scores, item-total correlations could serve as one index of item desirability, assuming that maximizing internal consistency of the scale scores is an operative goal.

Note, however, that some researchers have suggested that maximizing internal consistency is a misguided approach to scale development (Boyle, 1991; Cattell, 1973, p. 88; Epstein, 1983). Specifically, because items with the highest item-total correlations often share a considerable proportion of variance, subtests containing only these items probably will not maximize variance in the *original* test score. Also, by maximizing reliability, one may have created a "narrower" measure with suppressed validity coefficients. These issues highlight the importance of simultaneously considering multiple indices of item and scale quality when making scale-reduction decisions.

We developed a diverse set of such quality metrics, operationalized them in numeric form, factor analyzed them to understand the commonalities and differences among them, and used them to select a reduced set of items for each JDI subscale. A detailed description and literature review justifying this methodology appeared in Sinar et al. (1999). A brief overview of the techniques appears in the appendix. Next, we ascertained the psychometric and validity characteristics of scores from the resulting shortened scales, including a confirmatory factor analysis (CFA) to compare the network of relations of the long and short versions to external criteria. These analyses all occurred on a sample (designated below as Study 1) in which the full-length version of the scales (i.e., the full 72 items) had been administered. We also wanted to ensure that the shortened scales continued to exhibit desirable properties when administered in isolation from all the dropped items. Thus, as a second step, we collected an additional sample of data (designated below as Study 2) in which only the shortened versions of the scales were administered. This second sample served as a cross-validation of the validity and psychometric qualities of the abridged scales.

Study 1

Method

Participants and procedure. Participants comprised a national sample of 1,609 workers stratified by state population (55% male). We stratified the sample to obtain sufficient representation from states with lower populations and to oversample minority workers. Within stratifications, sampling was random. Responses resulted from a mailed distribution of 7,000 surveys, for a 23% response rate. No incentive was offered for survey completion, and all

surveys remained anonymous. The response rate we obtained compares favorably with other organizationally relevant research that used mail administration of paper-and-pencil survey instruments without direct participant incentives (Roth & BeVier, 1998).

Measures. The survey contained the five facet scales from the 1997 revision of the JDI: Work, Pay, Promotions, Supervision, and Coworkers (Balzer et al., 1997). In addition, we included a 6-item measure of intentions to quit the organization (ITQ) (Parra, 1995; based on Mobley, Horner, & Hollingsworth, 1978) and the 18-item JIG scale (Ironson et al., 1989). We scored all scale items (after reverse coding where necessary) in the standard JDI fashion with 0 for "no", 1 for "?", and 3 for "yes." To eliminate missing data, we first deleted all respondents with more than three missing data points across all JDI, ITQ, and JIG scales in accordance with the scoring instructions in Balzer et al. (1997). Next, we converted all remaining missing scale data points to scores of 1 (i.e., undecided). This procedure resulted in a final sample size of 1,534.

Results

Item quality indices. To operationalize our system of deciding on items to retain or drop, we developed a total of 10 different metrics of item quality. A full description of how each metric was generated appeared in Sinar et al. (1999), and an abbreviated description appears in the appendix. To illustrate one such process, we used the item response theory statistical analysis program, MULTILOG, to generate an *a* (discrimination) parameter for each item in the context of its complete facet scale. The *a* parameter quantifies the degree to which the item discriminates between levels of the "trait" (in this case a facet of job satisfaction) exhibited by the respondent (see Hambleton, Swaminathan, & Rogers, 1991). A low value of *a* indicates that an item contributes little to the scale in terms of discriminating between different levels of job satisfaction, whereas a high value indicates that the item provides satisfactory discrimination. With other concerns held constant, a higher level of *a* is usually preferable.

The 10 metrics of item quality were difficult to compare because many were on different scales and had very different means and standard deviations (see Table 1), so we standardized each metric by creating *z* scores in reference to all other items within a given facet scale. For example, in the Work facet, the item "bad" had the highest *a* parameter ($a = 2.84$) and thus also the highest standard score ($z = 2.44$) among the original 18 items in the Work facet scale. Next, we conducted a factor analysis of the 10 vectors of standard scores to ascertain common underlying dimensions. An exploratory factor analysis using a principal components extraction and varimax rotation yielded four

Table 1
Item Evaluation Criteria, Descriptive Statistics, and Methods of Measurement

Item Evaluation Criterion	*M*	*SD*	Range	Method of Measurement	Factor Analysis Loading
IRT *a* parameter	1.99	1.20	0.87-8.61	IRT *a* parameter using graded response model	Internal consistency
Factor saliency	0.61	0.11	0.37-0.87	Factor coefficient on first factor of exploratory analysis	Internal consistency
Item-total correlation	0.57	0.09	0.35-0.78	Corrected item-total correlation with scale score	Internal consistency
R^2 variance explained	0.07	0.15	0.00-0.70	Incremental variance proportion explained in regression of total scale score on individual item score	Internal consistency
Correlation with Intention to Quit (ITQ)	–0.29	0.07	–0.53-–0.16	Correlation of item with ITQ scale score	External validity
Correlation with Job in General (JIG)	0.36	0.10	0.18-0.65	Correlation of item with JIG scale score	External validity
Best job-worst job differential	43.43	16.33	9.00-86.00	A rating score from workers ($N = 104$) indicating the items most descriptive of workers' best or worst job experiences	Subjective quality
Expert rating	3.81	0.47	2.57-4.58	Professional Job Descriptive Index user ($N = 43$) rating of item quality on a 1 to 5 scale	Subjective quality
Percentage "don't know"	10.38	3.68	4.40-21.70	Percentage of respondents answering with "?" for item	Dispersion
Item extremity[a]	15.47	9.99	0.10-37.90	Absolute value of item endorsement rate minus 50%	Dispersion

a. Criterion reverse coded.

factors with eigenvalues in excess of one. A diagnostic scree plot also suggested the existence of four factors. The final column in Table 1 shows the factor with which each index was salient.

We interpreted the four factors as internal consistency, external correlations, subjective/user qualities, and dispersion. Inclusion of items with high scores on the internal consistency factor would tend to increase the Cronbach's alpha coefficient for the scale scores. Retaining items with high scores on the external correlations factor would tend to heighten the magnitude of the facet subscale's correlation with general job satisfaction and/or intentions to quit. Items with high scores on the subjective/user qualities factor were those that apparently had higher face validity as judged by a sample of expert users of the JDI and a sample of typical respondents. Items with high scores on the dispersion factor tended to have the lowest amount of response extremity. Put differently, low dispersion indicated high endorsement rates of one particular response option (e.g., the "yes" option). We calculated a factor score for each of these four factors by forming a unit-weighted average of the *z* scores of the individual indices salient with each factor. Thus, each item in all five of the facet scales (72 items total) had four quality metric values, each of which summarized a different aspect of item quality.

Item selection. At this point, we mixed empirical results with professional judgment to select a set of five items for each of the scales. Within each facet scale, we focused our attention on the items that, according to their quality metrics, contributed most to internal consistency, external correlations, subjective quality, and dispersion. Naturally, items that performed well on one metric did not necessarily perform well on the other metrics. Nonetheless, in most of the facet scales, there were a handful of items that performed well on all four metrics. Selecting among the high performing items required professional judgment calls. In particular, to avoid generating a response set, we wanted to maintain a roughly even split between positively and negatively worded items. For cases in which two or more of the best items appeared redundant in content, we looked farther down the list for nonredundant item content. Finally, given the brief nature of the new instrument, we wanted to avoid using the same adjective (e.g., "bad") in more than one facet scale. We selected five items for each scale with the expectation that we could maintain suitable internal consistency for scores from scales of that length. Experimentation with scales of different lengths indicated that alpha internal consistency reliability estimates dropped too low with fewer than five items. Results of our selection process appear in Table 2. Note that we selected three positively worded phrases and two negatively worded phrases for each facet subscale.

Table 2
Abridged Job Descriptive Index (JDI) Items With Corrected Item-Total Correlations From Study 1 and Study 2

		Corrected Item-Total Correlations	
JDI Facet Scale	Item Content	Study 1	Study 2
Work	1. Gives sense of accomplishment	.70	.63
Work	2. Dull	.69	.71
Work	3. Satisfying	.65	.69
Work	4. Uninteresting	.69	.61
Work	5. Challenging	.64	.58
Pay	1. Fair	.49	.66
Pay	2. Underpaid	.67	.68
Pay	3. Income adequate for normal expenses	.53	.42
Pay	4. Well paid	.63	.48
Pay	5. Insecure	.33	.34
Promotion	1. Good chance for promotion	.72	.72
Promotion	2. Dead-end job	.59	.61
Promotion	3. Promotion on ability	.63	.64
Promotion	4. Good opportunities for promotion	.68	.71
Promotion	5. Unfair promotion policy	.40	.37
Supervision	1. Praises good work	.57	.55
Supervision	2. Annoying	.58	.64
Supervision	3. Tactful	.59	.65
Supervision	4. Bad	.61	.74
Supervision	5. Up to date	.52	.55
Coworkers	1. Helpful	.62	.59
Coworkers	2. Boring	.55	.46
Coworkers	3. Intelligent	.65	.51
Coworkers	4. Lazy	.58	.49
Coworkers	5. Responsible	.65	.62

Note. JDI Items Copyright 1997, Bowling Green State University. Please contact first author for permission to use these scales.

Statistics for reduced scales. For all of the analyses that follow, we divided the Study 1 sample into two halves. The two subsamples were generated through a random case selection process that resulted in $n = 782$ for the subsample containing the full-length scales and $n = 752$ for the subsample containing the abridged scales. We chose this strategy to ensure independence in all of the statistical comparisons of Study 1 data.

The items shown in Table 2 were scored and summed using the standard JDI scoring system. Because of the reduced number of items in each scale, the abridged scale scores had a possible range of 0 to 15. Means and standard deviations for the abridged scales are shown in Table 3 alongside the values for the full-length scales. Note that the reduced item count in the abridged

scales resulted in much lower means and standard deviations for these scales (because fewer items contributed to the summated scores). Table 3 also shows the coefficient alpha reliability estimates for scores on all five facets before and after the scale reduction. As expected, because of the positive relationship between scale length and coefficient alpha, these values were all reduced for the abridged scales because of the smaller number of items in each scale. For all five abridged scales, however, alpha values were above the .70 threshold recommended by Nunnally and Bernstein (1994). All five abridged facet scales had substantial variability. Intercorrelations among the five abridged facet scales, JIG, and intentions to quit are shown above the diagonal in Table 3 with corresponding values for the full-length scales below the diagonal for comparison. Generally, correlations for the abridged scales were somewhat suppressed relative to the full-length version, as would be expected from the lower reliabilities of the abridged scale scores. Importantly, however, the general pattern of correlations appears to have been preserved. This finding provided initial evidence regarding the validity of scores on the abridged scales.

Table 3 also reports skewness and kurtosis statistics for full-length and abridged scales. As previous work would suggest (Balzer et al., 1997), all of the scales have distortions from normality even in the full-length versions. We were concerned that reducing the number of items in each facet scale might substantially exacerbate these distortions relative to the full-length versions. Using the respective standard errors of skewness (0.087) and kurtosis (0.175) from the smaller sample (a smaller error term and therefore a more conservative test), we calculated critical ratios of the differences in skewness and kurtosis between the full-length and abridged scales. Skewness worsened on the Work ($z = –4.1$, $p < .001$) and Coworker ($z = –2.6$, $p < .01$) facet scales and improved on the Promotions facet ($z = 3.1$, $p < .01$) as a result of reducing the length of the scales. Kurtosis worsened on the Promotions facet ($z = –2.9$, $p < .01$) as a result of reducing the length of the scale. The kurtosis and skewness parameters for the other facets remained essentially unchanged between the full-length and abridged versions of the scales.

Confirmatory comparison of correlation matrices. To ascertain more precisely the degree to which the correlation matrix changed between the abridged and full-length versions of the JDI subscales, we employed a technique conceptually similar to multigroup CFA. In CFA, one typically attempts to confirm a hypothesized pattern of correlations between a set of items (indicators) and one or more latent factors (unobserved or hypothetical constructs). In multigroup factor analysis, one seeks to confirm the hypothesized structure in multiple samples simultaneously. We used JDI subscale scores rather than items as indicators and focused on the equivalence of intercorrelations between the scales rather than a particular factor structure,

Table 3
Study 1 Descriptive Statistics and Scale Intercorrelations of Job Descriptive Index (JDI) Facets, Job in General (JIG), and Intentions to Quit (ITQ)

Scale	M^a	SD^a	Skewness[a]	Kurtosis[a]	M^b	SD^b	Skewness[b]	Kurtosis[b]	1	2	3	4	5	6	7
1. Work	37.25	13.59	–0.74	–0.47	11.16	4.99	–1.10	–0.16	(.86) .90	0.28	0.36	0.39	0.40	0.69	–0.55
2. Pay	14.70	8.47	–0.17	–1.17	8.21	5.05	–0.14	–1.27	.37	(.76) .86	0.33	0.24	0.23	0.40	–0.38
3. Promotions	8.77	7.98	0.90	–0.27	5.84	4.86	0.63	–0.77	0.41	0.36	(.81) .87	0.44	0.28	0.44	–0.38
4. Supervision	33.42	15.16	–0.40	–1.02	9.51	4.98	–0.42	–1.15	0.48	0.34	0.44	(.79) .91	0.45	0.52	–0.42
5. Coworkers	35.62	14.54	–0.62	–0.61	10.76	4.73	–0.85	–0.52	0.51	0.31	0.33	0.50	(.82) .91	0.51	–0.38
6. JIG	38.26	13.32	–0.99	0.27	38.77	13.08	–1.20	0.84	0.73	0.44	0.42	0.58	0.57	(.92) .92	–0.66
7. ITQ	1.07	1.10	0.55	–1.17	1.03	1.07	0.59	–1.07	–0.55	–0.45	–0.41	–0.50	–0.42	–0.66	(.90) .90

Note. Analyses of full-length JDI conducted on subsample with $n = 782$. Analyses of Abridged Job Descriptive Index (AJDI) conducted on subsample with $n = 752$. Full-length JDI correlations are presented below the diagonal with AJDI correlations above the diagonal. Coefficient alpha reliability estimates are presented on the diagonal with parenthetical values from the AJDI.

a. Full-length JDI scale.

b. Abridged (5 items) JDI scale.

but these considerations did not alter the essential CFA methodological approach or logic.

For Study 1, we compared the two correlation matrices that appear in Table 3. One of the underlying assumptions of multigroup structural equation modeling is independence of the samples (see Bollen, 1989, p. 356). As mentioned above, we randomly divided our sample to provide two independent subsamples to satisfy this assumption.

We began by testing a baseline model in which all correlations except one were allowed to vary between the two matrices. We made the assumption that the correlation between JIG and ITQ was identical between the two subsamples (because these scales were unchanged), so this correlation was fixed as equal across the two matrices. We then fit two progressively more restrictive nested models. The first model fixed all of the "internal" correlations—that is, the correlations among the five facets—to be identical across the two matrices. The second model additionally forced the correlations of each facet with JIG and ITQ—the external correlations—to be identical across the two matrices.

The top half of Table 4 details the results of these model tests. The difference between each nested pair of models is represented by a chi-square difference test. In general, a statistically significant value for a chi-square difference test would show that fit was worsened in the constrained model and therefore that the less restrictive model provided a better fit to the data. In contrast, a statistically nonsignificant chi-square difference test would indicate that the more restrictive model provided the best fit. Note, however, that the chi-square difference test is influenced by sample size (Bollen, 1989), which in this study was quite large. Thus, the chi-square difference tests reported here were likely to be sensitive to relatively small changes between the correlation matrices. Thus, in addition to considering the chi-square differences, one should examine changes in model fit indices to ascertain whether a restriction has substantially worsened model fit.

In our first model comparison, fixing the internal correlations among the facets did not cause a statistically significant increase in chi-square. This result indicated that the more restrictive model fit better, indicating that the pattern of correlations among the facets remained unchanged when the scales were shortened. As the next step in our model test, we fixed the "external" correlations of the JDI scales with JIG and ITQ. Fixing the external correlations did not cause a statistically significant increase in chi-square. This result suggested that the pattern of correlations between the facets and two external measures—JIG and ITQ—did not change between the full and abridged measures. Note that the general model fit indices (e.g., goodness-of-fit index) also showed excellent model fit for all three models. This result suggested that all observed differences between the two correlation matrices were relatively insubstantial.

Table 4
Confirmatory Comparisons of Correlation Matrices From Study 1 and Study 2

Model	Chi-Square	*df*	Δ Chi-Square (Δ*df*)	GFI	NNFI
Study 1 (original sample: full length vs. abridged)					
Baseline	0.0	1	—	1.00	1.01
Internal correlations fixed	10.8	11	10.8 (10)	1.00	1.00
All correlations fixed	18.3	21	7.5 (10)	1.00	1.00
Study 2 (new sample abridged vs. Original sample full length)					
Baseline	0.1	1	—	1.00	1.01
Internal correlations fixed	42.0**	11	41.9 (10)**	0.99	0.98
All correlations fixed	76.8**	21	34.8 (10)**	0.99	0.98
Study 2 (new sample abridged vs. Original sample abridged)					
Baseline	0.01	1	—	1.00	1.01
Internal correlations fixed	30.8*	11	30.7 (10)*	1.00	0.98
All correlations fixed	71.9**	21	40.1 (10)**	0.99	0.98

Note. GFI = goodness-of-fit index; NNFI = nonnormed fit index. Study 1 full version $n = 782$. Study 1 abridged version $n = 752$. Study 2 abridged version $n = 647$.
*$p < .01$. **$p < .001$.

Study 2

We conducted a second study in which the abridged JDI items were administered without the additional items from the full-length version of the instrument. We considered this an important step because we wanted to demonstrate the viability of using the abridged subscales by themselves. In other words, we wanted to ascertain whether context effects (i.e., the presence of the full list of items) substantially affected the psychometric characteristics of the subscales. Thus, the main analytic focus of Study 2 was a comparison of the correlation matrices for the abridged scales between the Study 1 sample and the Study 2 sample.

Method

Participants and procedure. Participants comprised a sample of 647 workers (33% male) from a large midwestern university. No faculty or student workers were included in the sample. Responses resulted from an initial distribution to the entire worker population of 1,200 surveys via campus mail, for a response rate of 54%. No incentive was offered for survey completion, and surveys were anonymous. Respondents returned completed surveys to a central receiving location (not in human resources). (Note that the complete set of abridged JDI facet subscales, the JIG scale, and our ITQ scale fit comfortably on one side of a sheet of 8.5-by-11-inch paper.) Pilot testing of

the survey indicated that the administration time of the abridged JDI facet subscales had been halved in comparison to the administration time of the full-length subscales.

Measures. The survey contained the five abridged JDI scales developed in Study 1. In addition, we included the same 6-item measure of ITQ (Parra, 1995) as used in Study 1 and the 18-item JIG scale (Ironson et al., 1989). We scored all scale items (after reverse coding where necessary) with 0 for each "no" response, 1 for each "?", and 3 for each "yes." After eliminating responses with missing data on the scales of interest using the identical strategy as in study (n = 11 responses dropped), the final sample size was 636.

Results

Descriptive statistics. Because of the reduced number of items in each abridged scale, scores had a possible range of 0 to 15. Means, standard deviations, and coefficient alpha reliability estimates for scores on all measures appear in Table 5. As expected, coefficient alpha reliability estimates for the abridged scales were similar to those obtained for the abridged scales in Study 1 and somewhat lower than the full-length scales in Study 1. For all five abridged scales, however, these values were still above Nunnally and Bernstein's (1994) recommended .70 threshold.

Note the skewness and kurtosis statistics reported in Table 5. In general, neither statistic was extreme for the abridged scales. The skewness (–1.69) and kurtosis (1.92) statistics for the Coworkers scale raised some concern about the distribution of scores on this facet scale. Specifically, the distribution was both negatively skewed and leptokurtic. Inspection of the frequency histogram indicated that a very large proportion of the respondents obtained the maximum possible score on the scale. This result may also explain the smaller standard deviation for the Coworker facet scores in comparison with the other facets.

Confirmatory comparison of correlation matrices. Using the abridged scales from Study 2 (correlation matrix in Table 5) and both the full-length and the abridged scales from Study 1 (see Table 3), we conducted similar multigroup comparisons to those reported in Study 1. These analyses showed the extent to which the scale intercorrelations from Study 2 replicated those in Study 1. Note that meta-analyses of job satisfaction in conjunction with other variables (e.g., Tett & Meyer, 1993) have consistently shown substantial between-samples variability in the correlations of job satisfaction with other constructs. Likewise, research on the facets of the JDI suggest that the correlations between facets can also vary from sample to sample (Balzer

Table 5
Study 2 Descriptive Statistics and Scale Intercorrelations of Job Descriptive Index Facets, Job in General (JIG), and Intentions to Quit (ITQ)

Scale	*M*	*SD*	Skewness	Kurtosis	1	2	3	4	5	6	7
1. Work	12.27	4.30	–1.54	1.16	0.84						
2. Pay	7.71	4.71	0.01	–1.24	0.23	0.75					
3. Promotions	6.02	4.97	0.55	–0.95	0.31	0.27	0.82				
4. Supervision	10.51	4.98	–0.81	–0.71	0.40	0.19	0.38	0.83			
5. Coworkers	12.80	3.58	–1.69	1.92	0.38	0.28	0.24	0.28	0.76		
6. JIG	41.85	11.50	–1.53	2.26	0.65	0.29	0.37	0.48	0.50	0.92	
7. ITQ	5.25	5.64	0.77	–0.62	–0.43	–0.23	–0.31	–0.42	–0.28	–0.64	0.90

Note. $N = 647$. Coefficient alpha reliability estimates are presented on the diagonal.

et al., 1997). Thus, these confirmatory analyses were undertaken with the expectation that some differences would appear among the matrices we compared.

In the baseline models, all correlations except one were allowed to vary between samples. We made the assumption that the correlation between JIG and ITQ would be unchanged between the two samples because the item content was identical to that in Study 1, so this value was fixed across samples. This assumption gave the models one degree of freedom, making them overidentified and enabling the calculation of global fit statistics. As in Study 1, we then fit progressively more restrictive nested models. The first models fixed all of the "internal" correlations—that is, the intercorrelations among the five facets—to be identical across two matrices. The second models additionally forced the correlations of each facet with JIG and ITQ (i.e., the external correlations) to be identical across the two matrices.

The lower two thirds of Table 4 details the results of the model tests. Fixing the internal correlations among the facets caused a small but nonetheless statistically significant increase in chi-square in comparisons with the full-length and abridged subsamples from Study 1. This result indicated that the baseline models fit better, suggesting a change in the pattern of correlations among the facets between the two samples. In the most restrictive models, fixing the external correlations also resulted in a statistically significant chi-square increase in both comparisons. This result suggested that the pattern of correlations between the facets and two external measures (JIG and ITQ) also changed between the samples. Note, however, that the general model fit indices were very high for all models, suggesting that observed differences between the pairs of correlation matrices were insubstantial.

General Discussion

The intention of the present studies was to use a systematic methodology to develop a brief or abridged version of the JDI facet measures of job satisfaction that preserved the desirable characteristics of this well-known and widely used measure. One of our component goals was to obtain a set of measures that exhibited similar patterns of correlations with external criteria. The criteria we used to anchor our comparisons were the JIG scale developed by Ironson et al. (1989) and a scale measuring intentions to quit the organization developed by Parra (1995). In addition, we wanted to carry forward the relationships among the facet scales themselves. Thus, in shortening the facet scales, we wanted to minimize changes to the correlative relations between each scale and other measures.

The confirmatory comparisons of correlation matrices from Study 1 and Study 2 suggest that we have substantially, though imperfectly, achieved this goal. Comparisons between full-length and abridged facets in Study 1 sug-

gested no distortions of correlative relationships as a result of shortening the facet scales. In contrast, the chi-square difference tests reported in comparisons with the Study 2 correlation matrix indicated that the intercorrelations among the facets and between the facets and external criteria changed to a modest degree relative to the matrices from Study 1. A certain degree of change, for example in the correlation between pay satisfaction and intentions to quit, is to be expected between samples. In particular, Study 1 was a cross section of the whole U.S. workforce, whereas Study 2 consisted of staff members from a university. Logically, one would not expect to reproduce identical validity evidence for job attitudes in two such divergent samples, and meta-analyses of job satisfaction support this claim (e.g., Tett & Meyer, 1993).

In addition to maintaining correlative relations with marker constructs in the nomological net for job satisfaction, we wished to maintain the high degree of internal psychometric quality that the full-length JDI has exhibited. Our two foci in this area were to maintain an acceptable level of internal consistency for each scale and avoid exacerbating undesirable distributional characteristics often observed in measures of job satisfaction. On the first goal we succeeded, with no alpha reliabilities below .75. On the second goal, we largely succeeded with the possible exception of the Coworkers facet scale. Future research efforts will have to experiment with alternative or additional items that reduce or prevent the ceiling effect observed in Study 2.

Finally, we wished to maintain the desirable user features that have been the hallmark of the full-length JDI. In particular, we wanted to maintain face validity, balance the use of negatively and positively worded items, and avoid redundant item content both within and between the facet scales. To achieve these ends, we employed both worker and expert ratings of JDI content. By selecting among the items with the highest ratings by workers and experts, we ensured that the most face-valid items were retained in the abridged scales. In addition, we used our own professional judgment to avoid redundant item content and keep a balance of negatively and positively worded items. Inspection of the item content presented in this article suggests that we largely succeeded on these goals. At the same time, we clearly met our goals of reducing the number of items in the subscales, the amount of area required on the survey instrument, and the administration time.

In summary, the abridged version of the JDI simultaneously preserves many desirable characteristics of the full-length version of the scale while reducing the item count, administration time, and required survey space for the instrument. The abridged instrument is suitable for modern multivariate organizational research. At the same time, we concur with Clark and Watson's (1995) sentiment that the work of scale development is never truly complete. We have identified several issues, such as the apparent ceiling effect in the Coworkers scale, that require additional study. Future research on the

abridged JDI should also assess its correlative and predictive characteristics in reference to a larger selection of external criteria such as organizational commitment and contextual performance. These efforts should also occur in a more diverse set of worker populations than presented here in Study 2. Finally, future research on scale development and maintenance should assess the applicability of our item metrics and item selection methodology to additional scales. Evidence from Sinar et al. (1999) and the present study suggests that a systematic strategy for assessing item quality can provide scale developers with useful information for making item retention decisions.

Appendix
Overview of Item Quality Index Generation Techniques

1. Item response theory (IRT) *a* parameter using graded response model: We used the graded response model provided by the software package MULTILOG (Thissen, 1991 **NEED REF**) to estimate scale item IRT *a* and *b* parameters for each of the five full-length facet scales.
2. Saliency with first factor of exploratory analysis: We ran principal components analysis with varimax rotation on each facet scale and recorded the factor pattern/structure coefficient of each item on the first component.
3. Corrected item-total correlation with scale score: We generated corrected item-total correlations between each item and its respective scale score.
4. Incremental variance proportion explained in regression of total scale score on individual item score: Using incremental variance explained as a criterion for entry, we ran regressions using the component items as predictors and the scale score as the criterion. We recorded the incremental variance explained as each item entered the regression.
5. Correlation of item with Intentions to Quit scale score: Treating intentions to quit as an important outcome correlated with job satisfaction, we recorded each item's zero-order correlation with this scale.
6. Correlation of item with Job in General scale score: Treating general job satisfaction as an important correlate to facets of job satisfaction, we recorded each item's zero-order correlation with scores on this scale.
7. Worker ($N = 104$) affirmation of item content as best job descriptor minus affirmation as worst job descriptor: We conducted a study of "face validity" in which a convenience sample of employed adults selected from among items in each scale to choose items most descriptive of their best and worst jobs.
8. Job Descriptive Index user ($N = 43$) rating of item quality on a 1 to 5 scale: We conducted a study of face validity in which a convenience sample of professional organizational researchers rated the subjective quality of each item on a 5-point scale.
9. Percentage of respondents answering with "?" for item: On the theory that high endorsement rates of the "?" option suggested an item less descriptive than one endorsed more frequently as "yes" or "no," we calculated the proportion of "?" responses within each item relative to "yes" and "no" responses.
10. Absolute value of item endorsement minus 50%: We calculated a measure of response extremity by discarding "?" responses and treating "N" as 0 and "Y" as 1. The item mean across all respondents is thus similar to the difficulty parameter *p* for dichotomously scored items and is a proxy for variance.

References

Balzer, W. K., Kihm, J. A., Smith, P. C., Irwin, J. L., Bachiochi, P. D., Robie, C., Sinar, E. F., & Parra, L. F. (1997). *User's manual for the Job Descriptive Index (JDI; 1997 Revision) and the Job in General (JIG) Scales*. Bowling Green, OH: Bowling Green State University.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, *12*, 291-294.

Buckley, M. R., Carraher, S. M., & Cote, J. A. (1992). Measurement issues concerning the use of inventories of job-satisfaction. *Educational and Psychological Measurement*, *52*, 529-543.

Cattell, R. B. (1973). *Personality and mood by questionnaire*. San Francisco: Jossey-Bass.

Clark, L. A., & Watson, D. (1995) Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). *The experience of work: A compendium and review of 249 measures and their use*. London: Academic Press.

Cranny, C. J., Smith, P. C., & Stone, E. F. (Eds.). (1992). *Job satisfaction*. New York: Lexington.

DeMeuse, K. P. (1985). A compendium of frequently used measures in industrial-organizational psychology. *The Industrial-Organizational Psychologist*, *23*, 53-59.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*, 360-392.

Ganzach, Y. (1998). Intelligence and job satisfaction. *Academy of Management Journal*, *41*, 526-539.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hoppock, R. (1935). *Job satisfaction*. New York: Harper and Brothers.

Ironson, G. H., Smith, P. C., Brannick, M. T., Gibson, W. M., & Paul, K. B. (1989). Construction of a Job in General scale: A comparison of global, composite and specific measures. *Journal of Applied Psychology*, *74*, 1-8.

Johnson, S. M., Smith, P. C., & Tucker, S. M. (1982). Response format of the Job Descriptive Index: Assessment of reliability and validity by the multi-trait, multi-method matrix. *Journal Applied Psychology*, *67*, 500-505.

Locke, E. A., Smith, P. C., Kendall, L. M., Hulin, C. L., & Miller, A. M. (1964). Convergent and discriminant validity for areas and rating methods of job satisfaction. *Journal of Applied Psychology*, *48*, 313-319.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. (1978). An evaluation of the precursors of hospital employee turnover. *Journal of Applied Psychology*, *63*, 408-414.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Organ, D. W. (1988). A restatement of the satisfaction-performance hypothesis. *Journal of Management*, *14*, 547-557.

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, *48*, 775-802.

Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology*, *77*, 963-974.

Parra, L. F. (1995). *Development of an intention to quit scale*. Unpublished manuscript, Bowling Green State University.

Rogelberg, S. G. (Chair). (1998, April). *Surveys and more surveys: Addressing and dealing with oversurveying*. Symposium conducted at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, Texas.

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science*, *7*, 60-65.

Roth, P. L., & BeVier, C. A. (1998). Response rates in HRM/OB survey research: Norms and correlates, 1990-1994. *Journal of Management*, *24*, 97-117.

Sinar, E. F., Julian, A. J., Stanton, J. M., Balzer, W. K., Smith, P. C., & Parra, L. F. (1999, April). *Strategic issues in reducing the length of psychological measures*. Poster presented at the annual meeting of the Society for Industrial and Organizational Psychology. **LOCATION?**

Smith, P. C., Balzer, W., Brannick, M., Chia, W., Eggleston, S., Gibson, W., Johnson, B., Josephson, H., Paul, K., Reilly, C., & Whalen, M. (1987). The revised JDI: A facelift for an old friend. *The Industrial-Organizational Psychologist*, *24*, 31-33.

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1969). *The measurement of satisfaction in work and retirement*. Chicago: Rand McNally.

Smith, P. C., Smith, O. W., & Rollo, J. (1974). Factor structure for Blacks and Whites of the Job Descriptive Index and its discriminations of job satisfactions among three groups of subjects. *Journal of Applied Psychology*, *59*, 99-100.

Smith, P. C., & Stanton, J. M. (1999). Perspectives on the measurement of job attitudes: The long view. *Human Resources Management Review.* **VOLUME & PAGES?**

Steers, R. M., Porter, L. W., & Bigley, G. A. (Eds.). (1996). *Motivation and leadership at work*. New York: McGraw-Hill.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on meta-analytic findings. *Personnel Psychology*, *46*, 259-293.

Zedeck, S. (1987, October). *Satisfaction in union members and their spouses*. Paper presented at The Job Satisfaction: Advances in Research and Practice Conference, Bowling Green, Ohio.

PLS. PROVIDE 4 OR 5 KEY WORDS FOR ELECTRONIC ACCESS TO YOUR ARTICLE