Construction of a Job in General Scale: A Comparison of Global, Composite, and Specific Measures

G. H. Ironson
Stanford University

P. C. Smith
Bowling Green State University and
Smith, Sandman, & McCreery, Perrysburg, Ohio

M. T. Brannick
University of South Florida

W. M. Gibson and K. B. Paul
Bowling Green State University

We describe the construction of a Job in General (JIG) scale, a global scale to accompany the facet scales of the Job Descriptive Index. We applied both traditional and item response theory procedures for item analysis to data from three large heterogeneous samples (N = 1,149, 3,566, and 4,490). Alpha was .91 and above for the resulting 18-item scale in successive samples. Convergent and discriminant validity and differential response to treatments were demonstrated. Global scales are contrasted with composite and with facet scales in psychological measurement. We show that global scales are not equivalent to summated facet scales. Both facet and global scales were useful in another organization (N = 648). Some principles are suggested for choosing specific (facet), composite, or global measures for practical and theoretical problems. The correlations between global and facet scales suggest that work may be the most important facet in relation to general job satisfaction.

Researchers have shown much interest in the topic of job satisfaction, demonstrated by the large number of articles (estimated by Locke, in 1976, at 3,350) written on the subject. It continues to be a major dependent variable in industrial, organizational, and social psychology. The Job Descriptive Index (JDI; Smith, Kendall, & Hulin, 1969, 1975/1985) has been reported to be the most frequently used measure of job satisfaction (De Meuse, 1986; O'Connor, Peters, & Gordon, 1978; Yeager, 1981). A computerized library search of *The Social Sciences Citation Index* and *Psychological Abstracts* revealed 454 articles referring to the JDI between January 1977 and November 1987. The focus of our article is to describe the development of a global Job in General scale (JIG) to accompany the JDI. The first purpose of this article is to address the question of why a global scale is needed and to investigate its psychometric properties, including item statistics, convergent and discriminant validity, differential response to treatment, and relation to the facet scales of the JDI.

A second purpose is to explore the usefulness of facet versus composite versus global scales. This issue, although explored in the context of job satisfaction, is clearly relevant to other areas of psychology as well.

Overview

Because these two purposes are intertwined and several studies are imbedded in this article, an overview of the organization of the article is in order. Much of the introduction is devoted to providing background for describing the usefulness of specific, composite, and global scales. This discussion lays the groundwork for the interpretation of data later in the report and helps to explain the need for a global scale.

We then discuss the development and validation of a new scale of satisfaction, the Job in General (JIG) scale, covering its construction, item statistics, reliability, and validity. In the discussion of discriminant validity, the distinctions among facet, composite, and global scales are again examined. Next, the JIG is examined in the context of an intervention. Finally, the Discussion section summarizes some of the salient issues.

Specific, Composite, and Global Scales

Facet Versus General Scales

This section describes what is meant by specific (facet) scales and general scales. *Facet* scales are intended to cover separately the principal areas within a more general domain. Each is intended to be relatively homogeneous and discriminably different from the others. The scales are, however, usually somewhat correlated with other scales in the same general domain. The long form of the Minnesota Satisfaction Questionnaire (MSQ; Weiss, Dawis, England, & Lofquist, 1967) lies close to the specific extreme of a specific–general continuum. It provides scores for 20 aspects of job satisfaction. The Quality of Employment

Survey (Quinn & Staines, 1979) measures six features of the job: Comfort, Challenge, Financial Rewards, Relations with Coworkers, Resource Adequacy, and Promotions.

The JDI (Smith et al., 1969, 1975/1985) covers five facets: Work, Pay, Promotions, Supervision, and Coworkers. Adjectives or brief phrases are presented; the respondent merely indicates whether or not each describes his or her job. Items refer to a single content area but differ in specificity. For example, on the Work scale, one quite specific item is *respected,* whereas others such as *good* are more general. One could easily construct facet scales with even more specific items grouped in even more specific groupings, for example, *quiet, adequate lighting,* and *rough toilet paper* as items in a working conditions scale. Facet scales are used to differentiate different aspects of job satisfaction; for example, to diagnose strengths and weaknesses in various sections of an organization.

In contrast, general scales are used to estimate the respondent's general overall feelings about the job. These feelings are expected to predict important behavior, such as quitting or being absent. They are widely used as indexes of organizational effectiveness.

Global Versus Composite Scales as General Measures

One approach to obtaining general measures is to ask directly about overall feelings about the job. Global scales ask the respondent to combine his or her reactions to various aspects of the job in a single integrated response. They assume that some sort of processing takes place and ask for its end product. During this process, the respondent may incorporate other aspects not measured in the facet scales or items. The respondent is asked something like "All things considered, how do you feel about your job most of the time?" Examples of multiple-item scales of this kind include those of Brayfield and Rothe (1951), Hackman and Oldham (1975), Hoppock (1935), and Quinn and Staines's (1979) facet-free scale. The Faces scale (Kunin, 1955) and the Adjectives scale (Ironson & Smith, 1981) are examples of single-item global scales with anchors.

Composite scales take a different approach to obtaining a general measure. They assume that the whole is equal to the sum of its principal parts. A general scale is thus formed by extensive coverage of components. An example is the short form of the MSQ (Weiss et al., 1967), which sums one item from each of the 20 scales of the long form. It has been shown, nevertheless, to be essentially unidimensional (Shiflett, Turney, & Cohen, 1979). Similarly, many researchers have summed the scores on five subscales of the JDI, despite serious problems with this procedure. The JDI scales were not designed to be summed. They were constructed to measure five discriminably different areas. Furthermore, they are only moderately correlated (.25 to .45; Smith et al., 1969) and represent at least five factors. Hence, the JDI scales cannot simply be assumed to measure a single unitary construct.

Composite measures, therefore, do not necessarily give the same results as asking for summary evaluations as in global scales (see Scarpello & Campbell, 1983; Schmidt & Kaplan, 1971; Smith et al., 1969). To summarize the arguments, composite scales may not be sufficient for estimating general satisfaction for several reasons:

1. Facet scales may omit some areas that may be important to the individual (Scarpello & Campbell, 1983). The composite may thus insufficiently estimate general job satisfaction.

2. Facet scales, similarly, may include items or areas that are relatively unimportant for a particular person.

3. Facet scales may have a descriptive as well as an evaluative component; they may not, therefore, fully reflect the general temperamental characteristics or affectivity of the individual.

4. The frame of reference for answering facet scales may differ from that of a global scale. The descriptive nature of the facet scales may, for example, elicit a more short-term response (Ryan & Smith, 1954).

5. Simply adding facets or combining them in a single linear manner for all people may not capture the unique individual method of combining components to arrive at a summary feeling. Global scales permit the respondent to do what comes naturally: to combine aspects of the situation as he or she ordinarily thinks of them.

Despite these objections, a good linear combination of the JDI facet scales might actually provide a good estimation of general satisfaction. Ferratt (1981) presented evidence that a linear combination of these facets did as well in predicting overall job satisfaction as any curvilinear or multiplicative model. The variety of content could even be an advantage in picking up valid general variance. (This is essentially the point made for criteria by Schmidt & Kaplan, 1971.) A combination of the items in the JDI subscales was therefore sought to form a composite scale.

Levels of Specificity

Measures at all levels of specificity have been widely used and found useful in both theoretical and practical research. Smith (1976) has argued for matching the specificity of a measure to the specificity of the criterion to be predicted.

But no one to date has determined whether approaches at both the global and the facet extremes are actually desirable, or whether the compromise of a composite measure is useful, and, if so, under what conditions each is preferable. In addressing this problem, a global companion scale to the JDI was constructed, focusing on the overall feeling about the job.

The scales would then be available to permit comparison of (a) separate facets, (b) a composite general scale, and (c) a global general scale.

Development of the JIG Scale

Overview of Studies

The development and validation of the JIG scale involved studies of several samples. Each study is described in the context of the type of information it furnishes about the JIG scale rather than in chronological sequence. To aid the reader who wishes to reconstruct the separate studies, the following samples were used:

1. Civil Service workers, Florida (N = 1,149)—item statistics, item response theory (IRT) analyses, and reliability.

2. Bowling Green archival samples (N = 3,866)—reliability.

3. Bowling Green archival sample (N = 4,490)—IRT analyses.

4. County employees, Florida (N = 227)—convergent validity and correlation of JIG and other general scales with JDI facets.

5. Nuclear power plant construction (N = 648)—discriminant valid-

ity, response to treatment (construct validity), and comparison of specific, composite, and global scales.

Construction

The JIG was developed to have the following characteristics: (a) multiple items to furnish an estimate of internal consistency; although testing stability over time does not require multiple items, it is meaningful only when the situation remains constant (Schneider & Dachler, 1978); (b) ease of reading and response, for use in working populations; (c) minimal overlap of content with measures of supposedly different variables; the global satisfaction measure should not, for example, describe job characteristics or ask about the intention to leave; (d) demonstrated convergent validity; and (e) compatibility with the JDI, because it was primarily intended to be used following the completion of the facet scales of the JDI.

The usual steps in careful construction of tests were followed in developing the JIG. First, we assembled a collection of 42 evaluative adjectives and short phrases concerning summary feelings about the job from a survey of the literature. Items were avoided that referred to facets or aspects of the job. An item such as *the chances for advancement on this job* (MSQ; Weiss et al., 1967) or *your fellow workers* (Warr, Cook, & Wall, 1979) would have been considered too specific for this scale. In addition, items were chosen that were evaluative and global rather than descriptive and specific (in contrast with the JDI) and had a long- rather than a short-term frame of reference. Thus, the JIG scale was intended to differ from the JDI in three important respects: more global, more evaluative, and longer in time frame.

The list of 42 items was given first to several samples of employees of an urban county in Florida. The combined samples yielded a total sample of 1,149. Twenty-two of the 42 items were negatively worded. The format and scoring were the same as those for the JDI (Smith et al., 1969; 1975/1985): *yes, no,* and *?*. The best items from this 42-item pool were tentatively selected by using several criteria:

1. High item-total correlations (all $rs > .45$, $Mdn = .65$).

2. High loading on the first principal component. A principal-components analysis with varimax rotation of all 42 items showed two clear factors; the unit weighted composites were moderately correlated. The first component was obviously a general factor. Typical items were *better than most, rotten, the pits,* and *acceptable*. This component accounted for 67.8% of the variance. The second component was unmistakably stress. The five items loading highly on this component were *stressful, tense, nerve-wracking, hectic,* and *pressured*.[1] This component accounted for 14.5% of the variance.

3. Adequately precise measurement throughout the satisfaction continuum. The percentage of respondents endorsing each item was one index of favorableness. In addition, 53 student judges rated each of the items for favorableness. Items were chosen to spread as evenly as possible across the range of favorableness while eliminating extremes. Furthermore, the IRT procedure (Hulin, Drasgow, & Parsons, 1983; Lord, 1980) was used to estimate difficulty and discrimination parameters for each potential item in the scale. The typical item analysis concentrating on internal consistency reliability favors selection of items in the middle of the scale (i.e., the middle of the scale for the particular calibration sample used for scale development). Item response theory, on the other hand, gives a better basis for choosing items so that precise measurement is available at all points for which discrimination is needed. An introductory discussion of latent trait theory as applied to attitude measurement may be found in Guion and Ironson (1983); see also Hulin et al. (1983).

The entire IRT procedure was duplicated for a second set of 4,490 subjects compiled from the Bowling Green data archives. They included pharmacists, hospital employees, health workers, retail food store employees, managers, employees in the construction of a nuclear power

Table 1
Item Statistics for the Job in General Scale

Item	Item-total r	% responding favorably[a]	a	b
			IRT parameters[b]	
Pleasant	.72	64	1.58	−.37
Bad[c]	.71	83	1.59	−1.19
Ideal	.52	28	.86	1.18
Waste of time[c]	.55	88	.90	−1.95
Good	.73	79	1.65	−.96
Undesirable[c]	.67	83	1.33	−1.28
Worthwhile	.67	85	1.34	−1.38
Worse than most[c]	.59	84	.86	−1.60
Acceptable	.68	80	1.24	−1.14
Superior	.48	23	.84	1.50
Better than most	.59	67	.76	−.69
Disagreeable[c]	.64	72	1.05	−.79
Makes me content	.59	49	.88	.19
Inadequate[c]	.66	68	1.15	−.58
Excellent	.59	32	1.34	.81
Rotten[c]	.58	85	.89	−1.67
Enjoyable	.74	69	1.61	−.54
Poor[c]	.67	82	1.27	−1.20

Note. Copyright 1985 by Bowling Green State University. The sample consisted of Civil Service workers in a large county in Florida, $N = 1,053$ (complete patterns).
[a] Favorable responses are *yes* to a positive item and *no* to a negative item. Proportion responding *yes* ranged from .05 to .85. [b] Item response theory (IRT) parameters: a = discrimination parameter; b = difficulty (favorability) parameter. [c] Reverse scored.

plant, engineers, word processors, and miscellaneous blue-, white-, and pink-collar employees.

Both sets of data were used to select evenly spaced and discriminating items, with primary emphasis on the more heterogeneous Bowling Green samples.

The resulting scale consisted of 18 global evaluative items. They are listed in Table 1 along with conventional item-total correlations, percentage answering favorably, and a and b parameters for the original Florida sample. Item totals exceed .47; favorability ranges from 23% to 88%. All a parameters (discrimination) are greater than .75; b parameters (difficulty or favorableness) range from −1.95 to +1.50.

Evaluation of Construction

Factor Structure

A principal-components factor analysis of the 18 selected items resulted (as expected) in one large factor, accounting for 87% of the variance. The correlation of scores derived from these 18 items with those from the total set of the original 42 items (even including the items that loaded on the second, stress factor) was .96.

Reliability

Internal consistency was checked first in the combined Florida sample ($N = 1,149$). Coefficient alpha was .91; considering

---

[1] These items, together with others, form the Stress in General scale (Ironson & Smith, 1978), to be used in conjunction with the Job Stress Index (Sandman & Smith, 1987) covering facets of job stress and overall job stress.

Table 2

Convergent Validity and Correlations of JDI Facets with the Job in General (JIG) Scale and Other General Scales

General measure	JIG	B-R[a]	F[b]	A[c]	N[d]
JIG	—				
Brayfield–Rothe (B–R)	.80	—			
Faces (F)	.75	.65	—		
Adjectival scale (A)	.76	.67	.81	—	
Numerical scale (N)	.67	.60	.75	.75	—
JDI					
Work	.78	.79	.68	.68	.59
Pay	.28	.22	.31	.30	.28
Promotions	.43	.38	.45	.40	.36
Supervision	.40	.21	.38	.34	.32
Coworkers	.42	.38	.32	.33	.33

Note. JDI = Job Descriptive Index. $N = 227$ county employees in Florida. All correlations are significant, $p < .01$.
[a] Brayfield–Rothe (1951). [b] Rating on scale of scowling to smiling faces (adapted from Kunin, 1955). [c] Rating scale with prescaled adjectives as anchors (Ironson & Smith, 1981). [d] Single item rating scale from −100 to +100.

the wide spread across the range of favorability, it is reasonably high. Alpha, in the Bowling Green samples each with $N > 100$, ranged from .91 to .95 (total $N = 3,566$). The information function (calculated using latent trait theory), which gives the approximate standard error of measurement at different levels of satisfaction, indicated precise measurement across the range.

Convergent Validity

Convergent validity was established by correlation with four other general scales of job satisfaction: the Brayfield–Rothe scale (1951); the Faces scale (Kunin, 1955); a rating scale anchored by adjectives prescaled for favorableness (Adjectives scale; Ironson & Smith, 1981); and a numerical rating scale (−100 to +100). The resulting correlations of the JIG ranged from .66 with the numerical scale to .80 with the Brayfield–Rothe scale, with the other correlations intermediate (see top of Table 2). Such levels are at least minimally acceptable (but certainly do not indicate equivalence of these four scales).

Evaluation of the JIG

Construct validity is the principal concern in evaluating the usefulness of the JIG. In this discussion we consider the interplay of theory and data in such areas as discriminant validity, differential response to treatment, and the relation between facet and global measures of satisfaction.

Discriminant Validity

Differences in validity encompass not only discriminant validity but also useful discriminant validity. The main goal in establishing discriminant validity was to determine whether there are differences between the JDI (either its component facets or their composite) and the JIG in (a) their patterns of correlations with other relevant variables or (b) their responses to situational changes.

This consideration really addresses a larger question of nonequivalence of measures. Is the newly developed JIG merely equivalent to the JDI? For two measures to be equivalent, they must meet certain statistical criteria. A long list of criteria was presented in Smith et al. (1969, pp. 153–156). Three particularly important criteria are (a) The two measures should be very highly correlated; (b) more stringently, they should show a similar pattern of correlations with a set of other variables; and (c) they should also respond in a similar way to treatments or to a change in the situation.

Discriminant validity would be demonstrated if the facet scales of the JDI tended to correlate more highly with specific measures than with general measures, while at the same time the JIG tended to correlate more highly with general than specific measures. The JDI scales, moreover, should correlate with relevant specific scales (Campbell & Fiske, 1959). Finally, differential response of the various scales to treatment could be examined as an indication of construct validity.

The Situation

An opportunity arose to test both the discriminant validity of the JIG and JDI scales and their differential response to treatment. The management of a nuclear power plant (under construction) reported low productivity and morale and sought assistance. On the basis of preliminary interviews, it was decided that supervision, communication, job definition, and trust were principal areas needing investigation. We assembled a pre-intervention battery of measures including the JDI, JIG, and a number of other scales measuring job characteristics and work attitudes.

As a result of this survey, a major intervention to improve feedback and job definition was undertaken. Working in small groups in weekly training sessions, supervisors practiced observing and recording incidents of behavior for each of their subordinates. The supervisors were to discuss the incidents with each subordinate within 48 hr of the observation. Supervisors then assigned these incidents to dimensions to form the basis for Behaviorally Anchored Rating Scales, or BARS (Smith & Kendall, 1963). The BARS scales were later used to provide feedback to the subordinates.

When the training sessions were being planned for intermediate levels of supervision, it became clear that another intervention had to be introduced. Very large groups of technical personnel proved to have only titular supervision. Lines of authority had to be redefined and clarified so that supervisory positions could be established. The persons promoted to these positions were then trained to observe and evaluate the next lower level of supervision. The use of behavioral observations could then be introduced throughout the rest of the organization.

Because of an impending change in top management, the intended postintervention survey had to be administered earlier than planned, when the intervention had reached only about two-thirds of the supervisory personnel. The resulting midintervention survey was expanded to include additional measures. The sample was large ($N = 648$). Therefore, there were data on hand for a comparison of the effectiveness of global,

composite, and specific kinds of scales in predicting a number of measures (described later).

Predictor Scales

Several measures were used as predictors of the series of dependent measures. These included five specific (JDI), one global (JIG), and one composite (theta) scale.

Specific scales (JDI). The set of five facet scales of the JDI. Each facet scale represents the sum of relatively specific descriptive items. For the Work Itself, Pay, Promotions, Supervision, and Coworkers scales, coefficients alpha for this sample were .78, .81, .87, .87, and .88, respectively.

Composite scale (theta). The best estimate of the presumed unidimensional variable underlying the JDI. Despite the low intercorrelations of the factors of the JDI, Parsons and Hulin (1982) showed that, in addition to the five usually found, a reliable unidimensional latent trait could be captured. In our study we followed their suggestion, using latent trait analysis, assuming a two-parameter model, and using LOGIST (Wood, Wingersky, & Lord, 1976) to estimate item and person parameters. The result seems roughly comparable with the first unrotated principal component extracted from the 72 items in the five scales, or with a second-order general factor. Presumably, as the best composite of items from the scales, it is a better general measure than the summed JDI. It seems to be largely descriptive and short-term, like the facets on which it is based.

Global scale (JIG). The sum of 18 global items in the same format as the JDI. This measure is global, evaluative, and relatively long-term. Coefficient alpha for this sample is .92.

Predicted (Dependent) Measures

There were 29 other measures on the midintervention test. Each was item analyzed and most were revised. All were scored so that high scores were favorable (except for Intent to Leave). To maintain their specific nature, we did not combine or group the specific scales in any way. Each was treated as a criterion and was predicted with a simultaneous multiple regression by using all five JDI scales, theta, and JIG. All 29 multiple correlations were statistically significant ($p < .01$). Those dependent measures with a multiple correlation below .25 were dropped so that only those explaining a meaningful percentage (> 6.25%) of the variance would be retained. Their elimination did not change the pattern of the results.

The first column of Table 3 shows the multiple correlations for the remaining 18 measures. They range from .29 to .70, indicating that the measures are presumably relevant and in the same general domain. The present article is not concerned with the particular scales used but with the pattern of results relevant to arguments of differential validity, that is, the pattern of global, composite, and specific predictors and criteria.

The measures that were included in the analysis were (a) Intent to Leave—3 items (adapted from Mobley, Horner, & Hollingsworth, 1978; alpha, this sample = .86); the Survey of Life Satisfaction (SOLS)[2]—18 items, JDI type (Ironson & Smith, 1978; alpha = .89); Trust in Management and Trust in Fellow Employees—two 6-item scales (Cook & Wall, 1980; alphas = .90 and .81); Identification with the Work Organization Index

(IWOI), Interest in Work Innovation Index (IWII), and Acceptance of Job Changes Index (AJCI)—2 to 3 items each (from Patchen, Pelz, & Allen, 1965; alphas = .50, .61, and .69); JCI: Skill Variety, Autonomy, Feedback, and Task Identity—4 to 6 items each (from the Job Characteristics Inventory; Sims, Szilagyi, & Keller, 1976; alphas = .60, .79, .81, and .82); Goal Clarity, Goal-Setting Feedback, and Goal-Setting Participation—3 to 4 items each (from the Goal-Setting Attributes scales of Arvey & Dewhirst, 1976; alphas = .81, .70, and .78); and Job Definition, Recognition, Merit System Needed, and Communication—3 to 4 items each (constructed for this situation; alphas = .73, .69, .54, and .60).

Intercorrelational Analyses of Predictors and Predicted Measures

First, how do the predictor scales correlate with each other? The primary question concerns the supposedly general scales: the composite and global scales. (Correlations with the five facet scales of the JDI will be discussed later.) Because of its frequent use by other researchers, we tentatively included a third possibility for a general measure in this part of the analysis. In addition to the JIG and the theta, we simply added all of the items from all of the five facet scales to create a summed JDI score.

The three general scales (theta, JIG, and summed JDI) formed a cluster. The highest correlation was between the scores from the summed JDI and theta (.921). (This is surprisingly large, because this summed JDI contained several less reliable items that were changed on the recent revision of JDI.[3] For further analyses, we decided to drop the summed JDI in favor of the unidimensional theta scale both because of their high intercorrelation and because theta is supposedly a more rigorous best estimate of a unidimensional construct underlying the JDI.) The JIG correlated only .665 with theta, and .660 with the summed JDI. These latter correlations are not high enough to suggest that either can be substituted for the other.

Second, how are the predicted measures related to each other? They are positively intercorrelated (except for the negatively worded Intent to Leave scale). There were no other significant negative relations among the 18 predicted measures,

---

Table 3

Comparison of Types of Scales for Ability to Predict

| Predicted measure | R^a | Predictor scales | | | | | |
| | | Specific (facet) | | Composite (theta) | | Global (JIG) | |
		r^b	$IR^{c,d}$	r^b	IR^c	r^b	IR^c
Intent to Leave	.56	—	—	—	—	−.54	.28
SOLS	.42	—	.10Pr	—	—	.37	.11
Trust in Managment	.61	—	.18Pr	—	—	.51	.10
Trust in Employees	.53	.52C	.23C	—	.07	—	—
IWOI	.51	—	.09C	—	.09	.46	.15
IWII	.33	.28W	.13C	—	.13	—	—
AJCI	.46	—	.10Pr	—	—	.42	.15
JCI							
Variety	.41	—	.08C	—	—	.39	.17
Autonomy	.42	—	—	—	—	.37	.11
Feedback	.58	.50S	.18S	—	—	—	.08
Task Identity	.42	.37W	.10P	—	—	.37	.12
Goal Setting							
Clarity	.61	.55S	.20S	—	—	—	.10
Feedback	.70	.65S	.33S	—	.06	—	—
Participation	.68	.63S	.28S	—	—	—	.12
Job Definition	.58	—	.12S	—	—	.50	.18
Recognition	.54	.44Pr	.20Pr	—	—	—	—
Merit System Needed	.37	.30P	.17P	—	—	—	—
Communication	.29	—	—	.24	—	—	—
Proportion^e		8.5/18		1/18		8.5/18	

Note. JIG = Job in General scale; SOLS = Survey of Life Satisfaction; IWOI = Identification with the Work Organization Index; IWII = Interest in Work Innovation Index; AJCI = Acceptance of Job Changes Index; JCI = Job Characteristics Inventory. Sample consisted of employees of a nuclear power plant under construction, $N = 648$. All entries are single or multiple correlations; all entries, $p < .05$.
^{a} Multiple correlation for five facets, theta, and JIG. ^{b} Highest bivariate correlation (W = work, P = pay, Pr = promotion, S = supervision, C = coworkers). ^{c} Significant ($p < .05$) incremental correlations after entry of six other independent variables. ^{d} Only the facet with the highest incremental correlation is reported. ^{e} Proportion of measures best predicted by scales.

and most were significantly positive. The correlations (excluding those with Intent to Leave) ranged from −.01 to +.62.

Third, are the global, composite, and specific measures equivalent in the sense that the patterns of correlations are closely similar? In other words, do they predict the 18 dependent measures similarly? Or, in contrast, do the different types of scales actually furnish additional information? The predictive ability of unidimensional theta and the JIG scales are compared with that of whichever facet scale proves to be most predictive of each dependent measure. Several questions need to be addressed to examine whether these measures are equivalent and whether the JIG adds any information:

1. Do the specific (facet) scales predict better than the other two types for this sample of employees and criterion measures? As the simplest test, are the bivariate correlations with many of the tests really higher for the JDI scales than for the JIG and theta scales? The answer is "yes." In the second column of Table 3, ("Specific, *r*"), a correlation coefficient and a code letter are entered for the facet whenever at least one of the facets shows a higher correlation with a dependent variable than does the theta or the JIG scale. Keep in mind that all the correlations appearing in Table 3 are significantly greater than zero. An entry represents a correlation that was higher for specific scales than for

either the composite or the global scale. Each facet shows at least one entry. There are entries in 8 of 18 rows, plus one tie. A facet, therefore, is showing the largest correlation in 8.5 out of 18 comparisons (shown as Proportion in the last row of Table 3).

This proportion is a function of the choice of dependent measures and is only an indication that the facet scales are frequently the best predictors. Simply counting the number of comparisons in which a particular measure shows the highest bivariate correlation with different measures biases the results in favor of facet scales for two reasons. First, the dependent measures are intercorrelated. For example, the four correlations counted for the Supervision facet involved Goal-Setting and Feedback scales that were substantially intercorrelated. Second, there are five chances for a correlation with one of the five facets to exceed that with the composite or global scale. This bias in favor of the facets seems acceptable here, because it is the usefulness of the global and composite scales that is most in question.

Moreover, a more stringent question is, does at least one of the facets add incrementally—after the four other facets plus theta, plus the JIG scales? For example, did adding the JDI Supervision scale significantly increase the multiple correlation

with a particular measure over and above that obtained from the other six predictors: Work, Pay, Promotions, Coworkers, theta, and JIG?

The answer is again "yes." There is an entry in the third column of Table 3 (under "Specific, *IR*") if at least one incremental correlation is significant. Only the facet with the highest incremental correlation is reported; the abbreviation identifies that facet.

The facet scales, as expected, perform well in predicting scores on other measures. Each contains information different from and in addition to that in either composite or global scales and different from that in the other facet scales.

2. Does theta predict better than the others, using the same criteria? Not really. The fourth column of Table 3 shows only one (low) bivariate correlation greater for this type of scale than for the best facet alone or for the JIG. The fifth column shows that for only four measures did the computation of the underlying unidimensional value gain any incremental predictive power beyond the five facets and the JIG. Of course, collinearity limits the extent that theta can improve prediction.

3. Is the JIG scale useful in prediction? The last two columns of Table 3 show that this scale is indeed useful. Its bivariate correlation is largest in 8 of 18 cases, with an additional tie. The incremental multiple correlations (*IRs*) in the last column are the strongest evidence of the unique contribution of the JIG. They are significant in 12 of the 18 cases. Thus, in its own right, JIG predicts even after both a multiple regression of specific (five facets) and a composite (theta) of JDI have been used.

4. Are the differences simply due to artifacts? Explanations in terms of reliability, distributions, self-report biases, format, social desirability, and the demographics of age, sex, tenure, education, and department were systematically ruled out. Differences might be smaller (or larger), however, with different predicted variables.

5. Do the differences really matter? Are the differences among scales large enough to be of practical significance? Among the seven scales, prediction varies greatly, as is shown by the *IR* columns in Table 3. In addition, the *IRs* are sufficiently different that investigators might be led to different conclusions if they substitute one type of measure for another.

6. Can discriminant validity be demonstrated? Is there a meaningful pattern concerning the content and specificity of the scales that predict particular dependent measures? For the facets, which are specific, the reasons are, very obviously, differences in subject matter. Satisfaction with Coworkers predicts, not astonishingly, Trust in Fellow Employees; Satisfaction with Pay predicts Merit (pay) System Needed; and Satisfaction with Supervision predicts several variables concerning feedback and goal setting.

JIG, however, clusters with other general measures: Intent to Leave, Life Satisfaction, Trust in Management, and Identification with the Work Organization. It also picks up some of the general factor in the JCI and other scales.

Discriminant validity has thus been demonstrated, because the JIG shows significantly greater validity than the JDI scales in predicting some variables. Conversely, the facet scales are more closely related to others. Moreover, the content of predictor and predicted measures seems to correspond. The pattern of hits in Table 3 is impressive.

Effects of Treatments

Fortunately, there is further evidence concerning construct validity and useful differences in measurement. In the nuclear power plant, the facet measures responded differently to the interventions than did the JIG. For about two-thirds of the employees, extended discussions of job behavior had taken place and lines of responsibility had been specified or clarified, with the purpose of improving supervision.

First, mean scores of JDI and JIG were compared from preintervention to midintervention. Means for the Supervision and Coworkers facets increased significantly; the increases were significantly greater than for the other facets or for JIG.

Furthermore, there was a control group of sorts. Because the intervention had reached only two-thirds of the employees, scores for these participants could be compared with those for the nonparticipants. Participants averaged significantly higher for Supervision satisfaction. However, for the other satisfaction scales, including JIG, treatment effects were significantly smaller, sometimes negative, and not significantly different from zero. Despite the limitation that employees had not been randomly assigned to treatments, the nature of these differences seems to support the distinction between specific and global scales. The lines of evidence converge from both correlations and treatments. The global measure developed here (JIG) is not equivalent to these facet measures (JDI) nor to some combination thereof. This nonequivalence is evident despite the use of the same method (and hence shared method variance) in the scales.

Importance as Indicated by the Relation Between Facet Satisfactions and Global Satisfactions

Exploration of the relation between global and facet satisfaction is of interest not only because it provides additional information on the equivalence question and on construct validity but also because it has substantive implications for the difficult topic of importance that has plagued psychologists for 40 years. A high correlation with the JIG may indicate saliency of the facet.

The first column at the bottom of Table 2 shows the correlations of the five facets of the JDI with the JIG scale. It is clear that the Work scale consistently shows the highest relation with general job satisfaction. This relation is not an artifact of method. It occurred not only in several of the samples using the JIG scale but also in a subset of the Florida sample for whom the Brayfield–Rothe and other rating scales were available (columns 2–5 at the bottom of Table 2). This consistent result strongly suggests that work is the most important facet in overall satisfaction with the jobs in these samples.

Discussion

The combination of these lines of evidence supports our point that useful measures of job satisfaction can be constructed that vary on the continuum from specific to general. But these measures are not equivalent. Each contributes unique and useful information. The specificity of the measure should match the specificity of the criterion, as advocated by Smith

(1976). For example, the facet scales responded differentially to treatment and would seem to be more useful for diagnosing high and low areas of satisfaction. So, if we may be forgiven a trite military analogy, "Use a rifle to hit the center of a target."

The global scales, on the other hand, predicted the general measures. We do need them for this purpose. But they did not respond (quickly) to a specific treatment: "Use a cannon to blast a large area."

The composite heterogeneous summated scale (theta) did not perform well here; "five shots from a rifle do better than the pellets from a shotgun shell." In the case of theta (our shotgun shell), no target was hit clearly. Therefore, "Use a rifle for small targets, a cannon for big ones, and avoid using a shotgun."

In future research, opportunities should be seized to track the long-term effects of interventions and of changing economic, organizational, and social conditions. Repeated measures do furnish a powerful tool for quality control of procedures affecting people. Such longitudinal studies should include both specific and global measures. In any research, measures should be chosen with appropriate specificity in mind.

References

Arvey, P., & Dewhirst, H. (1976). Goal setting attributes, personality variables, and job satisfaction. *Journal of Vocational Behavior, 9,* 179–189.

Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology, 35,* 307–311.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–104.

Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment, and personal need non-fulfillment. *Journal of Occupational Psychology, 53,* 39–52.

De Meuse, K. P. (1986). A compendium of frequently used measures in industrial/organizational psychology. *The Industrial-Organizational Psychologist, 23*(2), 53–59.

Ferratt, T. W. (1981). Overall job satisfaction: Is it a linear function of facet satisfaction? *Human Relations, 34,* 463–473.

Guion, R. M., & Ironson, G. H. (1983). Latent trait theory for organizational research. *Organizational Behavior and Human Performance, 31,* 54–87.

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology, 60,* 159–170.

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

Hulin, C. L., Drasgow, F. S., & Parsons, C. K. (1983). *Item response theory: Applications to psychological measurement.* Homewood, IL: Irwin.

Ironson, G. H., & Smith, P. C. (1978). *The Survey of Life Satisfaction (SOLS).* Bowling Green, OH: Department of Psychology, Bowling Green State University.

Ironson, G. H., & Smith, P. C. (1981). Anchors away—The stability of meaning when their location is changed. *Personnel Psychology, 34,* 249–262.

Kunin, T. (1955). The construction of a new type of attitude measure. *Personnel Psychology, 8,* 65–78.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297–1349). New York: Wiley.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Mobley, W. H., Horner, S. O., & Hollingsworth, A. T. (1978). An evalua-

tion of precursors of hospital employee turnover. *Journal of Applied Psychology, 63,* 408–414.

O'Connor, E. J., Peters, L. H., & Gordon, S. M. (1978). The measurement of job satisfaction: Current practices and future considerations. *Journal of Management, 4,* 17–26.

Parsons, C. K., & Hulin, C. L. (1982). An empirical comparison of item response theory and hierarchical factor analysis in applications to the measurement of job satisfaction. *Journal of Applied Psychology, 67,* 826–834.

Patchen, M., Pelz, D., & Allen, C. (1965). *Some questionnaire measures of employee motivation and morale.* Ann Arbor, MI: Institute for Social Research.

Quinn, R. P., & Staines, G. L. (1979). *The 1977 Quality of Employment Survey.* Ann Arbor, MI: Institute for Social Research.

Ryan, T. A., & Smith, P. C. (1954). *Principles of industrial psychology.* New York: Ronald.

Sandman, B. A., & Smith, P. C. (1987, October). Development of a measure of perceived job stress. In *Job satisfaction: Advances in research and applications.* Symposium conducted at Department of Psychology, Bowling Green State University, Bowling Green, OH.

Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts here? *Personnel Psychology, 36,* 577–600.

Schmidt, F. L., & Kaplan, L. B. (1971). Composite or multiple criteria: A review and resolution of the controversy. *Personnel Psychology, 24,* 419–434.

Schneider, B., & Dachler, P. H. (1978). A note on the stability of the Job Descriptive Index. *Journal of Applied Psychology, 63,* 650–653.

Shiflett, S., Turney, J. R., & Cohen, J. L. (1979). *Use of self-report technology in the development of an organizational action-research program.* (ARI Tech. Rep., No. 400). Alexandria, VA: U.S. Army Research Institute for the Behavioral Sciences.

Sims, H. P., Szilagyi, A. D., & Keller, R. T. (1976). The measurement of job characteristics. *Academy of Management Journal, 19,* 195–212.

Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), *Handbook of industrial psychology.* Chicago: Rand McNally.

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*(4), 31–33.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149–155.

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

Smith, P. C., Kendall, L. M., & Hulin, C. L. (1985). *The Job Descriptive Index* (Rev. ed.). Bowling Green, OH: Department of Psychology, Bowling Green State University. (Original work published 1975)

Warr, P. B., Cook, J., & Wall, T. D. (1979). Scales for the measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology, 52,* 129–148.

Weiss, D. J., Dawis, R. V., England, G. W., & Lofquist, L. H. (1967). *Manual for the Minnesota Satisfaction Questionnaire.* Minneapolis: Industrial Relations Center, University of Minnesota.

Wood, R. L., Wingersky, M. S., & Lord, R. G. (1976). *LOGIST—for estimating ability and item characteristic curve parameters* [Computer program]. Princeton, NJ: Educational Testing Service. (ETS RM 76-6, Modified).

Yeager, S. J. (1981). Dimensionality of the Job Descriptive Index. *Academy of Management Journal, 24,* 205–212.