

## Evaluating Weighted Models of Measuring Job Satisfaction: A Cinderella Story<sup>1</sup>

ROBERT P. QUINN AND THOMAS W. MANGIONE  
*Survey Research Center, The University of Michigan*

Data obtained from a national probability sample of 1533 American workers were used to assess the validity of ten methods of weighting job satisfaction ratings by importance ratings. The methods varied systematically in terms of the input they employed, scale treatment, and the types of mathematical operations used for weighting. Each method was evaluated with data from two subsamples of workers who differed in terms of the magnitude of their intraindividual correlations between satisfaction and importance ratings. The four criterion variables used to evaluate the weighting methods were measures of overall job satisfaction, job-related tension, likelihood of leaving one's present job, and mental health. The data not only failed to support the hypothesis that the validity of job satisfaction ratings may be increased by weighting them by importance ratings but indicated, on the contrary, that importance-weighting actually *reduced* the validity of satisfaction ratings. This reduction was attributed to several statistical problems, principally the failure of satisfaction and importance ratings to meet the demanding scaling assumptions of weighting models. An inconsistency was pointed out between the research activities necessary for providing an ideal test of a weighting model and those necessary for developing a generally useful measure of job satisfaction based on the model.

Objects that are most salient to an individual are also those most likely to elicit attitudinal responses. With regard to attitudes toward work, for example, it has been shown that those job facets that are most important to workers are those that also receive the most extreme positive or negative satisfaction ratings (Friedlander, 1965; Locke, 1961; Mobley & Locke, 1970). Conversely, facets of the job that are of little importance are characterized by more restricted ranges of attitudinal responses. Since workers differ in terms of the facets of the jobs that are important to them, it follows that the validity of estimates of overall

<sup>1</sup>This research was conducted under contract with the Employment Standards Administration of the U.S. Department of Labor. Researchers undertaking such projects under government sponsorship are encouraged to express their own judgment. Interpretations or viewpoints stated in this paper do not necessarily represent the official position or policy of the Department of Labor.

job satisfaction based upon satisfaction ratings of particular job facets (e.g., fringe benefits, security, competent supervision, etc.) may be improved if ratings of the importance of these same facets are somehow taken into account.

Most of those who have hitherto advocated weighting job satisfaction ratings by importance ratings in estimating overall job satisfaction have based their advocacy on theoretical rather than empirical grounds (Glennon *et al.*, 1960; England & Stein, 1961; Froehlich & Wolins, 1960; Morse, 1953; Porter, 1962; Vroom, 1964; Youngberg, Hedberg & Baxter, 1962). For example, Young, Hedberg, and Baxter claimed that their data made more "sense" when using importance and satisfaction measures together rather than satisfaction measures alone. They failed, however, to provide any external criterion with reference to which the supposed greater validity of the weighted measures could be assessed. Glennon *et al.* maintain that it is necessary to subtract satisfaction ratings from importance ratings, but they do not provide any substantiating evidence, and even the logic of their subtraction operation is questionable (Evans, 1969). England and Stein concluded that not taking importance into consideration means that scores on an attitude scale may "hide more than they reveal (p. 302)." This conclusion was based, however, on observed attitudinal differences between occupational groups. In spite of their advocacy of importance ratings England and Stein did not even employ importance measures in their study let alone any criteria for validating their attitudinal measures.

Systematic attempts to test the hypothesis that the use of importance ratings will increase the validity of satisfaction ratings have without exception failed to confirm the hypothesis (Decker, 1955; Ewen, 1969; Schaffer, 1953; Larson & Owens, 1965; Locke, 1961; Mikes & Hulin, 1968). Each of the latter studies employed basically the same design: (1) workers were asked to rate a number of job facets in terms of both their satisfaction with each facet and the importance to them of each; (2) a simple, unweighted estimate of job satisfaction was computed from the satisfaction ratings alone; (3) a weighted measure or measures using one or more methods of combining the satisfaction ratings with the importance ones were also computed; and (4) the weighted and unweighted measures were compared in terms of their associations with an external criterion. In all the studies cited this external criterion measure was a job satisfaction scale that made no reference to particular facets of the job, such as the General Motors "faces" test (Kunin, 1955). Mikes and Hulin used in addition an indirect indicator of job satisfaction, turnover from the job, as a criterion. None of these studies demonstrated that the importance-weighted measures investigated correlated

more highly with the criterion measure than did the unweighted measures. This does not mean, however, that the weighted measures were necessarily invalid but simply that they were *no better than* the more economical unweighted ones.

The present investigation, like those just cited, was intended principally to assess the relative validities of unweighted and importance-weighted methods of estimating overall job satisfaction. Although its design was basically similar to those of the previous investigations, it attempted to circumvent some of their methodological limitations and to untangle a number of related empirical matters which have tended to confound the issue of whether to weight or not to weight satisfaction ratings.

1. *Failure to consider the limiting effects of the correlation between satisfaction and importance ratings.* Where "weighting" satisfaction ratings involves multiplying them by importance ratings and the two sets of ratings are positively correlated, the weighting operation clearly adds little information to the satisfaction ratings. In the extreme case where the two sets of ratings are perfectly correlated, weighting amounts to no more than squaring the satisfaction ratings. Previous investigations have confined assessing the validity of weighting methods to samples of workers undifferentiated as to the magnitude of the intraindividual correlations between satisfaction and importance ratings. Their lack of success in demonstrating this validity may have resulted from weighted methods being superior to unweighted ones *only* among workers whose intraindividual correlations between satisfaction and importance ratings is low. For this reason the present study distinguished two subsamples of workers, those with near-zero intraindividual correlations between their satisfaction and importance ratings and those whose ratings were positively correlated. It was hypothesized that the advantages of weighted methods over unweighted ones would be greater in the former subsample.

2. *Restricted score ranges.* With a relatively homogeneous sample the variances of importance ratings may quite well be restricted. Weighting operations may as a result be correspondingly limited in their capacity to realize the full benefits of weighting. Most of the previous assessments of weighting job satisfaction measures were based upon workers selected from single companies and/or within limited ranges of occupations. The resulting curtailed distributions of importance ratings may as a result have contributed to the common failure to demonstrate any superiority of weighted methods. The present study, however, was based on a national probability sample of American workers. The range of observed importance ratings was, therefore, as great as any obtainable from the whole working population.

3. *Limited criterion measures.* With the exception of Mikes and Hulin

(1968) previous studies evaluating weighted job satisfaction measures used but a single criterion for evaluating these measures—an independent measure of overall job satisfaction. While this criterion is certainly the most appropriate one to use, findings based upon its use may be limited by the validity of the criterion itself. Moreover, its sole use ignores the possibility that while a weighted measure may not exhibit higher associations with overall job satisfaction than an unweighted one, the weighted measure may still be a better predictor of more remote criteria of or correlates of job satisfaction. For this reason the present study, in addition to using overall job satisfaction as a criterion, also employed three more remote criteria: feelings of tension associated with the job; the worker's estimate that he would seek a new job in the immediate future; and an index of mental health.

4. *Confounding operations.* The problem of weighting involves more than a simple decision to weight or not to weight. Embedded in this decision are three other types of choices involving: whether items or indices will be used as inputs to the construction of the measure; the sophistication of the treatment of the scale units of the input data; and the type of operation used to represent the weighting principle. Unlike previous investigations the present study examined the validities of weighted and unweighted job satisfaction measures within a design that systematically varied these three sets of choices. This design is presented in Table 1.

TABLE 1  
METHODS OF ESTIMATING OVERALL JOB SATISFACTION FROM COMBINATIONS  
OF SATISFACTION AND/OR IMPORTANCE RATINGS<sup>a</sup>

| Scale treatment                       | Methods which do not use importance ratings |             | Methods which weight satisfaction ratings by importance ratings |             |                                  |                                  |
|---------------------------------------|---|-------------|---|-------------|----------------------------------|----------------------------------|
|                                       |   |             | Multiplicative model  |             | Ipsative model                   |                                  |
|                                       | Item input                                  | Index input | Item input  | Index input | Item input                       | Index input                      |
| Unit weighting                        | A   | B           | C   | D           | E <sub>1</sub><br>E <sub>2</sub> | F <sub>1</sub><br>F <sub>2</sub> |
| Normalized ( <i>z</i> ) scores        | G   | H           | I   | J           | —                                | —                                |
| Weights based on multiple regressions | K   | L           | M   | N           | —                                | —                                |

Note—Blank cells indicate logically inconsistent combinations of ipsatization and other methods. Methods implied by these blank cells were not tested.

<sup>a</sup> Each of the alphabetic entries in the body of the table corresponds to a method of estimating overall job satisfaction described in the Procedures section.

The first type of decision incorporated in the design involves the type of input used to derive the satisfaction measure from workers' ratings of job facets. On one hand, computations may be based directly on a total set of items. On the other hand, indices may be constructed from these items and computations may be based on the index scores rather than on the original items. When items are used as inputs the basic dimensions of job satisfaction that are represented by the largest number of items may make the greatest contribution to the total satisfaction scores. Where derived indices are used to represent these dimensions, employing indices as inputs may provide a more equitable weighting of the separate dimensions. Half of the weighting methods investigated in the present study used item inputs (Methods A, C, E, G, I, K, M in Table 1) and the remainder used index inputs (Methods B, D, F, H, J, L, N).

Most job satisfaction measures in use today employ unit weights in the computation of job satisfaction scores (e.g., ratings on a four-point integer scale where 1 may indicate "not at all satisfied" and 4 may equal "very satisfied"). Although such weights are sometimes referred to as "weighting items equally," this is in fact not the case. The *effective* weight of an item is dependent upon both its variance and its covariances with other items. The item (or index) with the greatest variance and smallest covariances will make the greatest contribution to the total score. Most models for weighting job satisfaction by importance ratings assume a *tabula rasa* situation in which all satisfaction items (or indices) have initially equal weights to which the importance ratings are then applied. Since this situation seldom, if ever, exists the simultaneous use of importance ratings and *unit* weights provides an unfair test of the utility of importance ratings. For this reason, the present study evaluated the use of importance ratings at three levels of scale treatment that differed in the extent to which they approximated the *tabula rasa* situation that was presumed to provide the best test of importance-weighted models of job satisfaction: using unit weights on a four-point integer scale (Methods A-F); using  $z$  scores which normalized each importance and/or satisfaction rating with reference to the total sample's mean and standard deviation (Methods G-J); and assigning scoring weights on the basis of multiple regression equations that used an independent measure of overall job satisfaction as a criterion (Methods K-N). The first of these made no provision for unequal item (or index) variances or covariances; the second gave all items (or indices) in any computation equal variances, and the third circumvented the problem of items (and indices) having different variances *and* covariances.

Two different types of importance-weighting methods were employed,

multiplication, and ipsatization. The former involved multiplying each worker's satisfaction ratings by his importance ratings that employed the same type of input and scale treatment (Methods C-D, I-J, M-N). The two "ipsative" methods (Methods E and F) restricted the computation of each worker's overall satisfaction score to those facets of the job that he had rated as most important.

## METHOD

Data were obtained from a 1969 survey of American workers conducted by the Survey Research Center of The University of Michigan. The principal aims of the survey were to determine the major problems faced by workers, to develop economical measures of job satisfaction suitable for use with samples of workers in heterogeneous occupations, and to assess the effects of working conditions on workers' job satisfaction and mental health. Detailed descriptions of the survey's sample and measures as well as basic univariate and bivariate tables are presented in Quinn *et al.* (1971).

### *Sample*

Personal interviews were conducted with a national probability sample of 1533 persons who were living in households, were 16 years old or older, and were working for pay for 20 hours a week or more. Since interviews were obtained from all eligible workers in a household, each worker had an equal probability of being selected, and the sample as a result was self-weighting.

### *Measures*

1. *Facet-specific job satisfaction items.* Twenty-three evaluative statements (e.g., "the pay is good") were rated at the end of the interview on a four-point scale in terms of how true the worker felt they were of his job.

2. *Facet-specific importance items.* The same 23 statements were rated at the beginning of the interview on a four-point scale under the instructions, "People differ a lot in terms of which of these things are more important to them. We'd like to know how important to *you* each of these things is." Approximately an hour elapsed between the worker's importance ratings and their satisfaction ratings.

3. *Facet-specific job satisfaction indices.* A factor analysis of the importance ratings of the 23 statements indicated that they could be represented by five orthogonal dimensions (Quinn *et al.*, 1971). For

each worker a set of five job satisfaction index scores were constructed that represented these five factors. Each index score was the mean of the salient items on the factor; the original four-point rating scales were used in computing means.

4. *Facet-specific importance indices.* A parallel set of five indices was constructed using the 23 importance ratings as inputs.

5. *Facet-free job satisfaction.* This index estimated each worker's overall job satisfaction from his responses to five questions that in no way referred to specific facets of his job: "All in all, how satisfied would you say you are with your job?" "If a good friend of yours told you (he/she) was interested in a job like yours for your employer, what would you tell (him/her)?" "Knowing what you know now, if you had to decide all over again whether to take the job you now have, what would you decide?" "In general how well would you say that your job measures up to the sort of job you wanted when you took it?" "I'd like to get some idea of the kind of job you'd *most* like to have. If you were free to go into any type of job you wanted, what would your choice be?" The worker was presented with fixed-alternative response categories for the first four questions. The fifth was coded in terms of whether or not he chose his present job.

6. *Job-related tension*, an eight-item version of the scale originally used by Kahn *et al.* (1964).

7. *Likelihood of leaving present job*, based upon answers to the question, "Taking everything into consideration, how likely is it that you will make a genuine effort to find a new job with another employer within the next year?"

8. *Negative mental health.* This 21-item measure was based upon a factor analysis of mental health items previously used by Lipman *et al.* (1967), Gurin *et al.* (1960), and Hunt *et al.* (1969). Its three major components were depression, physiological complaints, and performance debilitation.

9. *Intraindividual correlation between satisfaction and importance ratings.* For each worker a Pearson  $r$  was computed between his ratings of the 23 job facets in terms of his satisfaction with them (Measure 1) and his ratings of the same 23 job facets in terms of their importance to him (Measure 2). Two groups of workers were distinguished in the analysis: those with near-zero intraindividual correlations, defined as those with correlations ranging from  $-.40$  through  $.40$ , and those with positive correlations, defined as those with correlations ranging from  $.41$  through  $1.00$ . The class limits of these two groups were defined with reference to the  $.05$  confidence limits of the correlation coefficient, the degrees of freedom of which were 21.

*Procedure*

Sixteen different combinations of types of job satisfaction and/or importance input data (Measures 1-4), scale treatments, and weighting operations were used in order to predict workers' overall job satisfaction (Measure 5) and other reactions to their work (Measures 6-8). These 16 combinations, the alphabetic designations of which correspond to those in Table 1, were as follows.

A. *Unit weighting of items, not considering importance.* A mean<sup>2</sup> was computed of each worker's satisfaction ratings of the 23 job facets (Measure 1).

B. *Unit weighting of indices, not considering importance.* A mean was computed of each worker's satisfaction ratings on each of the five summary job satisfaction indices (Measure 3).

C. *Unit weighting of items, multiplication.* For each of the 23 job facets a worker's satisfaction rating on the original four-point scale (Measure 1) was multiplied by his importance rating on the original four-point scale (Measure 2). A mean of these 23 cross-products was then computed.

D. *Unit weighting of indices, multiplication.* For each of the five indices a worker's satisfaction score (Measure 3) was multiplied by his importance rating on the comparable importance index (Measure 4). A mean of these five cross-products was computed.

E. *Unit weighting of items, ipsatization.* A mean was computed of the worker's satisfaction ratings of job facets (Measure 1) *only* on those job facets which he had rated as "very important" to him (Measure 2). Workers could as a result differ considerably in terms of the number of satisfaction ratings that entered into their resultant overall job satisfaction score. A worker who rated all 23 facets as very important would have an overall satisfaction score based on all 23 of his job satisfaction ratings. A worker who rated only a few of the 23 facets as important would have an overall satisfaction score based on far fewer satisfaction ratings. Two versions of this index were, therefore, computed. The first required that the worker had rated at least 14 of the 23 items as "very important," and the second required that he had rated at least seven items as "very important."

<sup>2</sup>Obtaining a mean would have been identical to summing the input scores had there been no missing data on any of the input items or indices. Since many workers had missing data on at least one item a mean was used instead and had the effect of prorating each worker's valid ratings over the items that he had failed to rate. Any worker with a missing data score on a third or more of the input variables to any of the sixteen methods was excluded from the analysis of that method.



*F. Unit weighting of indices, ipsatization.* Two methods employing this principal were used:

1. A worker's overall satisfaction score was equated with his score on the satisfaction index (Measure 3) to which he assigned the highest importance rating (Measure 4). No estimate was made where a worker had two or more indices with identical "most important" ratings.

2. A worker's overall satisfaction score was the mean of his scores on the two satisfaction indices (Measure 3) to which he assigned the two highest importance ratings (Measure 4). No estimate was made where a worker had three or more indices with identical "most important" ratings.

*G. Normalized item scores, not considering importance.* Each of the 23 satisfaction items (Measure 1) was converted to  $z$  scores. A mean of these 23  $z$  scores was then computed.

*H. Normalized index scores, not considering importance.* Each of the five satisfaction indices (Measure 3) was converted to  $z$  scores. A mean of these five  $z$  scores was then computed.

*I. Normalized item scores, multiplication.* Each of the 23 satisfaction items (Measure 1) and importance items (Measure 2) was converted to  $z$  scores. After a constant had been added to remove negative signs, each satisfaction  $z$  score was multiplied by its corresponding importance  $z$  score. A mean of these 23 cross-products was then computed.

*J. Normalized index scores, multiplication.* Each of the five satisfaction indices (Measure 3) and importance indices (Measure 4) was converted to  $z$  scores. After a constant had been added to remove negative signs, each satisfaction  $z$  score was multiplied by its corresponding importance  $z$  score. A mean of these five cross-products was then computed.

*K. Multiple regression<sup>3</sup> employing items, not considering importance.* The 23 facet-specific satisfaction items (Measure 1) were used in a multiple regression analysis employing facet-free job satisfaction (Measure 5) as the criterion variable. Resulting beta weights were then assigned to each of the items, and the mean of the items thus weighted was computed.

*L. Multiple regression employing indices, not considering importance.* The five facet-specific satisfaction indices (Measure 3) were used in a multiple regression analysis employing facet-free job satisfaction (Measure 5) as the criterion variable. Resulting beta weights were then assigned to each of the indices, and the mean of the indices thus weighted was computed.

<sup>3</sup> In all four methods employing multiple regression, beta weights were obtained on a random half-sample of workers. Subsequent validation analyses were confined to the remaining half-sample.

TABLE 2  
 PRODUCT-MOMENT CORRELATIONS BETWEEN (1) SIXTEEN METHODS OF USING IMPORTANCE AND SATISFACTION RATINGS  
 TO PREDICT OVERALL JOB SATISFACTION AND (2) OVERALL JOB SATISFACTION  
 AND OTHER RESPONSES TO THE JOB

| Method <sup>a</sup>  | N <sup>b</sup>   | Facet-free<br>job<br>satisfaction | Job-related<br>tension | Likelihood<br>of leaving<br>present job | Negative<br>mental<br>health |
|--|------------------|-----------------------------------|------------------------|---|------------------------------|
| Methods which do not use importance ratings                        |                  |                                   |                        |   |                              |
| A Unit weighting of items  | 1502             | 46                                | 37                     | 29                                      | 20                           |
| B Unit weighting of indices  | 1485             | 45                                | 35                     | 29                                      | 19                           |
| G Normalized item scores   | 1502             | 46                                | 39                     | 28                                      | 20                           |
| H Normalized index scores  | 1485             | 45                                | 37                     | 28                                      | 19                           |
| K Multiple regression employing items                              | 756 <sup>c</sup> | 55                                | 21                     | 29                                      | 15                           |
| L Multiple regression employing indices                            | 756 <sup>c</sup> | 50                                | 33                     | 28                                      | 19                           |
| Methods which weight satisfaction ratings by<br>importance ratings |                  |                                   |                        |   |                              |
| C Unit weighting of items, multiplication                          | 1486             | 34                                | 28                     | 23                                      | 10                           |
| D Unit weighting of indices, multiplication                        | 1467             | 33                                | 26                     | 23                                      | 10                           |

|    |   |                  |    |    |    |    |
|----|---|------------------|----|----|----|----|
| I  | Normalized item scores, multiplication  | 1486             | 32 | 27 | 21 | 10 |
| J  | Normalized index scores, multiplication   | 1467             | 31 | 29 | 22 | 10 |
| M  | Multiple regression, items, multiplication  | 741 <sup>c</sup> | 39 | 26 | 18 | 10 |
| N  | Multiple regression, indices, multiplication  | 735 <sup>c</sup> | 32 | 27 | 15 | 07 |
| E1 | Unit weighting of items, ipsatization, with minimum of 14 "very important" items      | 417              | 52 | 45 | 32 | 17 |
| E2 | Unit weighting of items, ipsatization, with minimum of seven "very important" items   | 1060             | 51 | 42 | 32 | 21 |
| F1 | Unit weighting of indices, ipsatization, with no tied "most important" ranks          | 812              | 35 | 17 | 25 | 12 |
| F2 | Unit weighting of indices, ipsatization, with no more than two "most important" ranks | 956              | 42 | 31 | 32 | 19 |

Note—Decimal points and signs omitted. All significant correlations were in the predicted direction.

<sup>a</sup> Alphabetic designations of methods correspond to those in Table 1 and in the Procedures section.

<sup>b</sup> *N*'s vary slightly between columns due to missing data on the column variables.

<sup>c</sup> Reduced *N* is due to analysis being conducted on a random half-sample of workers.

*M. Multiple regression employing items, multiplication.* The 23 cross-products of satisfaction and importance items generated in Method I were used in a multiple regression analysis employing facet-free job satisfaction (Measure 5) as the criterion variable. Resulting beta weights were then assigned to each of the cross-products, and the mean of the cross-products thus weighted was computed.

*N. Multiple regression employing indices, multiplication.* The five cross-products of satisfaction and importance indices generated in Method J were used in a multiple regression analysis employing facet-free job satisfaction (Measure 5) as the criterion variable. Resulting beta weights were then assigned to each of the cross-products, and the mean of the cross-products thus weighted was computed.

## RESULTS

The first column of correlation coefficients in Table 2 shows the associations between overall job satisfaction as estimated by the facet-free measure and as estimated by various combinations of satisfaction and/or importance ratings of specific job facets. None of the methods in the first six rows of the table employed importance ratings of job facets. All six were based exclusively upon satisfaction ratings of facets and differed solely in the type of input used (items or indices) and scale treatment (unit weights,  $z$  scores, or weights based on multiple regression equations). The remaining ten rows of the table represent methods which, although differing in terms of type of input and scale treatment, had in common their weighting of satisfaction ratings by importance ratings either through multiplication (Methods C, D, I, J, M, N) or ipsatization (Methods E and F).

The method that should be regarded as the reference point for evaluating all correlations in Table 2 is Method A, a straightforward mean of each worker's satisfaction ratings of the 23 job facets. The method made no use of importance ratings, employed a simple four-point rating scale, and did not capitalize on the possible advantages of grouping the 23 items into five summary indices. It was the least sophisticated of all 16 measures and made the greatest number of questionable assumptions. In spite of Method A's unimpressive credentials, virtually none of the other 15 methods was a significantly<sup>4</sup> better predictor of overall job satisfaction. The single exception to this was Method K which, ignoring importance ratings, assigned weights to each of the 23 satisfaction ratings on the basis of a multiple regression analysis predicting

<sup>4</sup> Statistical significance was in all instances based on  $t$ -tests of differences between Pearson  $r$ 's using two-tailed tests.

facet-free job satisfaction from these same 23 ratings. That Method K was superior to Method A in predicting overall job satisfaction was hardly surprising inasmuch as the former's scoring weights were "custom fitted" to predict the criterion in question.

None of the ten measures employing importance ratings was superior to Method A in estimating overall job satisfaction. On the contrary, seven of the ten measures that used importance ratings provided significantly *poorer* estimates than Method A. Even the importance-weighted method that had the highest correlation with facet-free job satisfaction, Method E1 with its  $r$  of .52, was not significantly better than Method A and was, moreover, based on an ipsatization technique that provided an estimate of overall satisfaction for less than a third of the total sample. A second ipsative method, Method E2 with its  $r$  of .51, provided an estimate for only two-thirds of the sample. The two "best" importance-weighted methods suffered, therefore, from their inability to estimate the overall job satisfaction of a large percentage of workers.

In six pairs of methods (A/C, B/D, G/I, H/J, K/M, L/N) both members of each pair were comparable in terms of type of input and scale treatment but differed only in that the first method made no use of importance ratings while the second one multiplied satisfaction scores by importance scores. According to Table 2, all six methods that employed importance ratings to weight satisfaction ratings provided significantly poorer estimates of overall job satisfaction than did the comparable methods that used satisfaction ratings alone.

Table 2 further shows that in estimating overall job satisfaction it made little difference whether the input variables were the undifferentiated set of 23 ratings of job facets or the five index scores based on a factor analysis of these job facets. Among the methods that used importance weightings there was, in addition, no significant difference between methods in terms of the sophistication of their scale treatments.  $z$  scores offered no improvement over unit weighting, and weights based on multiple regression analyses in order to eliminate the effects of item covariances were no better than either unit weights or  $z$  scores. Among the methods that made *no* use of importance ratings there were no significant differences between methods based on unit weights and those based on  $z$  scores.

These estimates of the validity of various methods of combining satisfaction and/or importance ratings in order to estimate overall job satisfaction were limited by their exclusive concentration upon facet-free job satisfaction as a criterion. Since facet-free job satisfaction may itself have been an imperfect estimator of overall job satisfaction, three other

criterion measures were also used. Among the six methods that did not use importance ratings none of the more complex methods (i.e., those using indices rather than items or those using  $z$  scores or weights based on multiple regressions rather than unit weights) was superior to the more primitive Method A in predicting any of these three criteria (Table 2). The apparent advantage of weights based on multiple regression equations that was evident with regard to estimating facet-free job satisfaction disappeared where the three more remote criteria were concerned, and with regard to predicting job-related tension these weights were a significant handicap.

When the ten methods that used importance ratings were compared to Method A in terms of their correlations with job-related tension, likelihood of leaving present job, and negative mental health, there was no evidence justifying the superiority of the importance-weighted methods. Among the 30 comparisons made (Table 2) there were no instances in which a correlation involving an importance-weighted method was significantly greater than the correlation between Method A and the same more "remote" criterion measure. Comparing the six pairs of methods (Methods A/C, B/D, G/I, H/J, K/M, L/N) that were similar in terms of their types of input and scale treatment but differed only in the first method made no use of importance ratings while the second one multiplied satisfaction scores by importance scores, Table 2 further shows importance-weighting to be a considerable detriment in estimating job satisfaction. Among the 18 comparisons thus made in predicting the three more "remote" criteria, six comparisons showed that there was no difference between importance-weighted methods and otherwise comparable unweighted methods. The remaining 12 comparisons all indicated that the methods using importance ratings were significantly *poorer* predictors of job-related tension, likelihood of leaving present job, and negative mental health than were the comparable methods that completely ignored importance ratings.

It was hypothesized above that methods of estimating overall job satisfaction by multiplying satisfaction ratings of job facets by importance ratings would be superior to methods that ignored importance ratings principally where the intraindividual correlations between satisfaction and importance ratings were low. Where intraindividual correlations were high, weighting by importance would add little information to satisfaction ratings and, hence, would not improve upon methods based upon satisfaction ratings alone. To test this hypothesis, Table 3 shows the correlations between the 16 methods of estimating overall job satisfaction and the four criterion variables. Each correlation was computed separately for those workers who exhibited a significant

positive intraindividual association between importance and satisfaction ratings and those who had statistically near-zero intraindividual associations. The correlations in the last ten rows of the table, those that present data from the ten importance-weighted estimates of overall job satisfaction would seem to confirm the hypothesis that such weighted measures would be superior to unweighted ones principally among workers who had near-zero correlations between their importance and satisfaction ratings. Thirty-seven of these 40 correlations were at least marginally greater among the workers whose importance and satisfaction ratings were not associated than among those whose ratings were positively associated. Thirteen of these differences between correlation coefficients were statistically significant. This apparent confirmation of the hypothesis was, however, seriously undermined by the analogous correlation coefficients in the first six lines of Table 3—correlations based upon methods of estimating job satisfaction that did *not* use importance ratings and, hence, were assumed to be unaffected by intraindividual correlations between satisfaction and importance ratings. Fifteen of these 24 correlations were significantly greater among the workers whose importance and satisfaction ratings were not associated than among those whose ratings were positively associated. The hypothesis that the concurrent validity of estimates of job satisfaction that weighted satisfaction ratings by importance ratings would be greater among workers who had a near-zero correlation between their satisfaction and importance was, therefore, negated by the fact that among these same workers even estimates of job satisfaction that did *not* consider importance ratings appeared to be more valid than among other workers.

#### CONCLUSION

This study provided no empirical support for the hypothesis that the validity of job satisfaction measures can be improved through weighting satisfaction ratings by importance ratings. In failing to confirm this hypothesis the study's findings were entirely consistent with the few previous studies that had tested it. The previous tests had, however, suffered from several important limitations that were to a considerable extent overcome in the present study: (1) their use of restricted samples, a circumstance that might have resulted in restricted ranges of importance ratings; (2) their tendency to compare importance-weighted methods with unweighted methods while failing to hold constant such related factors as the type of input used (i.e., whether items or derived indices were used) and how scale units were assigned (i.e., through arbitrary scoring weights, through normalized scores, or through use of such empirically-based scoring weights as those obtained from multiple

TABLE 3  
 PRODUCT-MOMENT CORRELATIONS BETWEEN (1) SIXTEEN METHODS OF USING IMPORTANCE AND SATISFACTION RATINGS  
 TO PREDICT OVERALL JOB SATISFACTION AND (2) OVERALL JOB SATISFACTION AND OTHER RESPONSES TO THE  
 JOB—CONTROLLING ON INTRAINDIVIDUAL IMPORTANCE X SATISFACTION CORRELATION

| Method <sup>a</sup>   | Facet-free job satisfaction                  |   | Job-related tension                          |   | Likelihood of leaving present job            |   | Negative mental health                       |   |
|---|--|---|--|---|--|---|--|---|
|   | Near-zero<br>$r_{1 \times 8}$<br>$N = 980^b$ | Positive<br>$r_{1 \times 8}$<br>$N = 407^b$ | Near-zero<br>$r_{1 \times 8}$<br>$N = 981^b$ | Positive<br>$r_{1 \times 8}$<br>$N = 409^b$ | Near-zero<br>$r_{1 \times 8}$<br>$N = 856^b$ | Positive<br>$r_{1 \times 8}$<br>$N = 346^b$ | Near-zero<br>$r_{1 \times 8}$<br>$N = 966^b$ | Positive<br>$r_{1 \times 8}$<br>$N = 406^b$ |
| Methods which do not use importance ratings                     |  |   |  |   |  |   |  |   |
| A   | 47   | 35  | 41   | 24  | 32   | 17  | 20   | 14  |
| B   | 46   | 36  | 38   | 23  | 31   | 19  | 19   | 14  |
| G   | 47   | 34  | 43   | 25  | 32   | 16  | 19   | 14  |
| H   | 46   | 35  | 40   | 24  | 31   | 17  | 19   | 12  |
| K   | 53   | 54  | 20   | 10  | 37   | 19  | 10   | 17  |
| L   | 51   | 37  | 34   | 20  | 39   | 15  | 14   | 16  |
| Methods which weight satisfaction ratings by importance ratings |  |   |  |   |  |   |  |   |
| C   | 85   | 25  | 28   | 24  | 27   | 08  | 11   | 07  |
| D   | 83   | 26  | 25   | 24  | 27   | 10  | 11   | 07  |



|    |  |    |    |    |    |    |    |    |    |
|----|--|----|----|----|----|----|----|----|----|
| I  | Normalized item scores, multiplication   | 33 | 24 | 27 | 23 | 25 | 08 | 11 | 07 |
| J  | Normalized index scores, multiplication  | 32 | 25 | 26 | 23 | 26 | 09 | 11 | 07 |
| M  | Multiple regression, items, multiplication <sup>c</sup>  | 37 | 35 | 23 | 20 | 28 | 10 | 07 | 12 |
| N  | Multiple regression, indices, multiplication <sup>c</sup>  | 33 | 22 | 27 | 14 | 26 | 03 | 07 | 03 |
| E1 | Unit weighting of items, ipsatization, with minimum of 14 "very important" items <sup>d</sup>      | 49 | 34 | 44 | 17 | 34 | 21 | 17 | 11 |
| E2 | Unit weighting of items, ipsatization, with minimum of seven "very important" items <sup>e</sup>   | 49 | 39 | 44 | 24 | 35 | 15 | 20 | 18 |
| F1 | Unit weighting of indices, ipsatization, with no tied "most important" ranks <sup>f</sup>          | 33 | 22 | 17 | 10 | 20 | 23 | 10 | 05 |
| F2 | Unit weighting of indices, ipsatization, with no more than two "most important" ranks <sup>g</sup> | 41 | 34 | 34 | 16 | 30 | 30 | 21 | 04 |

Note—Decimal points and signs omitted. All significant correlations were in the predicted direction.

<sup>a</sup> Alphabetic designations of methods correspond to those in Table 1 and in the Procedures section.

<sup>b</sup> *N*'s vary slightly between columns due to missing data on the column variables.

<sup>c</sup> *N* for this correlation is roughly half that of that indicated in the column heading due to analysis being conducted on a random half-sample of workers.

<sup>d</sup> *N*'s are 264 (near-zero  $r_{1 \times s}$ ) and 93 (positive  $r_{1 \times s}$ ).

<sup>e</sup> *N*'s are 692 (near-zero  $r_{1 \times s}$ ) and 286 (positive  $r_{1 \times s}$ ).

<sup>f</sup> *N*'s are 543 (near-zero  $r_{1 \times s}$ ) and 246 (positive  $r_{1 \times s}$ ).

<sup>g</sup> *N*'s are 639 (near-zero  $r_{1 \times s}$ ) and 291 (positive  $r_{1 \times s}$ ).

regression analyses); (3) their almost exclusive concentration upon measures of overall job satisfaction as validating criterion, eschewing secondary and more remote criteria; and (4) their failure to take into account intraindividual correlations between satisfaction and importance ratings as a possible influence on the validity of importance-weighted methods.

The present study, substantially unencumbered by these limitations, not only replicated the unanimously "negative" findings of earlier validation studies but showed in addition that importance-weighting actually *decreased* the validity of job satisfaction ratings. The presumed theoretical "help" provided by the principle of importance-weighting proved instead to be an empirical hindrance in estimating overall job satisfaction. Weighting of satisfaction ratings by importance ratings, which earlier studies had shown to be at worst a rather innocuous theoretical indulgence, was shown by the present data to be something worse than innocuous.

The data produced, therefore, a statistical Cinderella story. The best method of estimating overall job satisfaction was clearly a naive method that was far less conceptually sound and mathematically ornate than its more elegant step-sisters. This does not necessarily mean that importance weighting completely invalidated satisfaction ratings. Even the worst of the weighted measures in Tables 2 and 3 still maintained significant correlations with facet-free job satisfaction. The best that can be said of importance-weighted measures of job satisfaction is that they are certainly no better than unweighted measures, but that if a researcher insists on using importance weights on theoretical grounds he should be aware of the fact that their use may not *too* greatly undermine the validity of his job satisfaction measure.

There are a number of possible explanations why importance-weighted job satisfaction measures were no improvement over unweighted ones. The first such explanation concerns matters relevant to the study's unique methodology, especially the possible limitations of its measures. The least important issue seems to be the validity of the criterion measures since these, whatever their possible limitations, were at least constant in all comparisons. As summarized in Method A, the ratings of the 23 job facets exhibited a high internal consistency as a job satisfaction measure, with a reliability of .88. With regard to the validity of Method A, additional analysis indicated that it was highly related to independent measures of the quality of workers' jobs (multiple  $R = .60$ ), and based upon data from a sample of 311 workers in heterogeneous occupations was significantly related to occupational turnover within a two-year period. There was less evidence for the more critical matter of

the validity of the importance ratings as measures of individual differences. All that is known at present is that they generated a replicable factor structure (Cobb & Quinn, 1971) and as descriptors of the relative rankings of what U.S. workers wanted from their jobs correlated .70 with rankings obtained by Herzberg *et al.* (1957) from entirely different sources of data (Quinn, 1971).

Another explanation is that inasmuch as the importance ratings and satisfaction ratings were positively correlated, weighting by importance added little new information to the satisfaction ratings. But the data indicated that weighted measures still proved inferior to unweighted ones even among workers whose intraindividual correlations between the two sets of ratings were near to zero. The ideal condition for testing the validity of weighting models would, of course, involve having two sets of ratings that were completely independent. This ideal situation may, however, be a wholly artificial one. Importance ratings have been shown to be associated with satisfaction ones, and a longitudinal study by Rosenberg (1957) suggested that the association may be more one of cause-and-effect than a methodological artifact. A dilemma therefore exists with regard to the best conditions for evaluating any weighting model—be this model multiplicative or ipsative, such as used in the present study, subtractive (Beer, 1966; Kuhlen, 1963; Pelz & Andrews, 1966; Porter, 1962; Ross & Zander, 1957; Spitzer, 1964),<sup>5</sup> or one even more complex (Evans, 1969; Spitzer, 1964). The best condition is that where there are near-zero intraindividual correlations among such panels of variables as satisfaction ratings of job facets, importance ratings, and ratings of levels of aspiration and attainment with regard to these facets. Where three sets of ratings are involved (as in Evans' use of importance, aspiration, and attainment), this condition may exist for relatively few people in any sample. As a result, while the model may be adequately tested with this restricted sample, an inadequate opportunity is provided for building a better measure of job satisfaction through the application of the model. Scoring weights based on a restricted number of cases may prove quite inadequate for the

<sup>5</sup> Evans (1969) has argued compellingly that subtractive models make no sense when only importance and satisfaction ratings are involved and should be confined to data indicating how much a worker wants of something and how much he receives. In spite of this, a *sub rosa* analysis was conducted in the present study using one method of combining importance and satisfaction ratings that subtracted the two. This analysis was confined to item input and unit weights. The results indicated that even this logically unjustified and virtually meaningless operation produced an estimate of overall job satisfaction that was certainly no worse than those based on the more logically sound (given the two types of ratings involved) principles of multiplication and ipsatization.

remainder—which may be the majority—of the sample. The ideal conditions for *testing* a weighting model may be quite different from those under which an improved and generally applicable measure of job satisfaction may be developed through invoking the model.

The average worker apparently is one step ahead of the model builders. He is often unobliging enough to engage in his own weighting procedures *before* he provides what the researcher presumes to be “pure” satisfaction ratings. There is, moreover, no available way to guarantee that a worker will respond to a satisfaction item without first applying the mental calculus that takes into account the saliency to him of that item. Ironically, the best course followed by a researcher interested in developing a measure of job satisfaction may be to encourage rather than discourage a worker to engage in such internal deliberations prior to rating job satisfaction items. This encouragement would, of course, provide a poor testing ground for any weighting model. But developing a measure of a variable and testing a model designed to predict the variable are not always compatible activities. Future research employing weighted models of job satisfaction could profit from a clearer distinction (admittedly blurred in the present investigation) as to whether the primary concern of the investigation is in instrument development or model testing.

A number of mathematical considerations may also have resulted in the present study’s demonstration of the uselessness of importance ratings. The operationalization of the multiplicative and ipsative models hewed quite closely to the mathematical principles implied in the models. Unfortunately, most proponents of such models are terribly vague in the methods they would regard as adequate operationalizations. The most abused term seems to be “equal weights.” The study attempted, however, to do full justice to each model by testing it under three different conditions of “equality,” using arbitrary scoring weights, normalized scores, and item weights based upon multiple regression equations. A further mathematical limitation of the models tested lay not in their one proposes a “weighted” model it is all too easy to assume that this ambiguity but in the naiveté of their mathematical specifications. When weighting just means simple linear combination of the terms involved. Other higher-order forms of combination are available and may conceivably provide better approximations of psychological realities. Hullian theory and many other multiplicative models have perhaps overaccustomed many psychologists to equating “weighting” with “multiplication.” When applied to job satisfaction the multiplication operation is certainly intuitively appealing, parsimonious, easily performed, and, given the primitive state of our present knowledge, without serious, more sophis-

ticated competitors. It may also be inappropriate. Its appropriateness is, however, an issue that only future research can decide. Meanwhile, its major statistical limitation is the demanding scaling assumption that it makes. Few, if any, measures relevant to job satisfaction (i.e., ratings of satisfaction, importance, level of aspiration, or level of attainment) even employ interval scales. Models of job satisfaction which involve subtracting attainment ratings from aspiration ratings are fairly conservative in assuming only interval scale data. But multiplicative operations go one step further by demanding nonexistent *ratio* scales. At present, ratings of facets of jobs in terms of their importance to the worker or his satisfaction with them have far less than ratio scale properties, and the application of multiplicative models to these ratings may do little but reduce the validity of the satisfaction ratings, a theoretically discouraging state of affairs which was nevertheless clearly demonstrated in the present study.

The study's implications for the application of more "sophisticated" models in the construction of job satisfaction measures are dismal. For lack of ratio scales, or even interval ones, most available weighting models cannot be adequately tested. When, for example, importance ratings of job facets are used to multiply the differences between goal aspirations and goal attainment with regard to these same facets (as suggested by Evans, 1969) theory will have far outrun available technology and may have greatly abused the three sets of ratings employed in the analysis. Such elaborate combinations may consequently produce little more than statistical garbage. Given the present inadequacy of rating scales, any demonstration in the near future of the validity of more elaborate weighting models is likely to be illusory. Probably it will have capitalized fortuitously on the correlations among the predictor ratings used in the model and the independent correlation each set of ratings has with whatever criterion is used to validate the model. The latter correlations offer some insurance that unadorned and unweighted satisfaction ratings of job facets may, as indicators of overall job satisfaction, survive whatever unjustified mathematical machinations may be inflicted upon them in the overzealous application of particular models.

Most future tests of such models, as well as most past tests, must overcome a still further limitation that not only existed in the present study but which might have been a major factor contributing to the study's failure to affirm the utility of importance-weighted measures of job satisfaction. This limitation hinges upon the distinction made above between the ideal conditions for testing a job satisfaction model and those conditions most appropriate to developing a good measure of job

satisfaction. In a study designed to test a weighting model, the selection of job facets to be rated in terms of their importance should reflect the model tester's desire to make the greatest distinction possible between important and unimportant facets of the job. As a result, the worker should be required to rate many job facets that most workers would regard as trivial. Unfortunately, present research constraints provide little opportunity to ask a worker to contemplate trivialities. If a researcher intends to ask workers to indicate their satisfaction with regard to a number of job facets in order to estimate their overall satisfaction, the facets he selects to be rated will generally be those he decides are relevant to most workers in the population he is studying. The present study was no exception. The 23 job facets that were rated by workers in terms of importance and satisfaction were selected so as *not* to be trivial. This selection resulted in each worker being asked to respond only to those job facets that the study's several pretests indicated had been salient for most of the workers in the sample. A worker's importance ratings may as a result have reflected little more than his marginal distinctions among job facets that by-and-large were *all* of considerable importance to him. The testing of importance-weighting models should not, however, rely upon marginal distinctions.

#### REFERENCES

- BEER, M. *Leadership, employee needs, and motivation*. Columbus, OH: Ohio State University, Bureau of Business Research, 1966.
- COBB, W., & QUINN, R. P. What workers want: Factor analyses of importance ratings of job facets. Ann Arbor, MI: Survey Research Center, 1971. Mimeo.
- DECKER, R. L. A study of three specific problems in the measurement and interpretation of employee attitudes. *Psychological Monographs*, 1955, **69**, 16. Whole No. 401.
- ENGLAND, G. W., & STEIN, C. I. The occupational reference group—A neglected concept in employee attitude studies. *Personnel Psychology*, 1961, **14**, 299-304.
- EVANS, M. G. Conceptual and operational problems in the measurement of various aspects of job satisfaction. *Journal of Applied Psychology*, 1969, **53**, 93-101.
- EWEN, R. B. Weighting components of job satisfaction (Doctoral Dissertation, University of Illinois). Ann Arbor, MI: University Microfilms, 66-4172, 1969.
- FRIEDLANDER, F. Relationships between the importance and the satisfaction of various environmental factors. *Journal of Applied Psychology*, 1965, **49**, 160-164.
- FROELICH, H. P., & WOLINS, L. Job satisfaction as need satisfaction. *Personnel Psychology*, 1969, **13**, 407-420.
- GLENNON, J. R., OWENS, W. A., JR., SMITH, W. J., & ALBRIGHT, L. E. New dimension in measuring morale. *Harvard Business Review*, 1960, **38**, 106-107.
- GURIN, G., VEROFF, J., & FELD, S. *Americans view their mental health*. New York: Basic Books, 1960.
- HERZBERG, F., MAUSNER, B., PETERSEN, R. O., & CAPWELL, D. F. *Job Attitudes: Review of research and opinion*. Pittsburgh, PA: Psychological Service of Pittsburgh, 1957.

- HUNT, S. M., SINGER, K., & COBB, S. Components of depression identified from a self-rating depression inventory for survey use. *Archives of General Psychiatry*, 1967, **16**, 441-447.
- KAHN, R. L., WOLFE, D. M., QUINN, R. P., SNOEK, J. D., & ROSENTHAL, R. A. *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley, 1964.
- KUNIN, T. The construction of a new type of altitude measure. *Personnel Psychology*, 1955, **8**, 65-77.
- KUHLEN, R. G. Needs, perceived need satisfaction opportunities, and satisfaction with occupation. *Journal of Applied Psychology*, 1963, **47**, 56-64.
- LARSEN, J. M., JR., & OWENS, W. A., JR. The assignment of job attitude items to subscales. *Journal of Applied Psychology*, 1965, **49**, 172-181.
- LIPMAN, R., LOVE, L., RICKLES, K., ULENHUTH, E., & LAZER, R. Selected measures of changes in outpatient drug evaluation. Paper delivered at A.C.N.P. meeting. San Juan, PR, 1967.
- LOCKE, E. A. Importance and satisfaction in several job areas. Paper delivered at American Psychological Association. New York, 1961.
- MIKES, P. S., & HULIN, C. L. Use of importance as a weighting component of job satisfaction. *Journal of Applied Psychology*, 1968, **52**, 394-398.
- MOBLEY, W. H., & LOCKE, E. A. The relationship of value importance to satisfaction. *Organizational Behavior and Human Performance*, 1970, **5**, 463-483.
- MORSE, N. C. *Satisfactions in the white-collar job*. Ann Arbor, MI: Survey Research Center, 1953.
- PELZ, D. C., & ANDREWS, F. M. *Scientists in organizations: Productive climates for research and development*. New York: Wiley, 1966.
- PORTER, L. W. Job Attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, **46**, 375-384.
- QUINN, R. P. What workers want: General descriptive statistics and demographic correlates. Ann Arbor, MI: Survey Research Center, 1971. Mimeo.
- QUINN, R. P., SEASHORE, S., MANGIONE, T., CAMPBELL, D., STAINES, G., & McCULLOUGH, M. *Survey of working conditions: Final report on univariate and bivariate tables*. Washington, D.C.: U.S. Government Printing Office, 1971. Document 2916-0001.
- ROSENBERG, M. *Occupations and Values*. New York: Free Press of Glencoe, 1957.
- ROSS, I. C., & ZANDER, A. F. Need satisfaction and employee turnover. *Personnel Psychology*, 1957, **10**, 327-338.
- SCHAFFER, R. H. Job satisfaction as related to need satisfaction at work. *Psychological Monographs*, 1953, **67**, No. 14, Whole No. 364.
- SPITZER, M. E. Goal-attainment, job satisfaction and behavior. (Doctoral dissertation, New York University). Ann Arbor, MI: University Microfilms, No. 64-10, 10,048, 1964.
- VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.
- YOUNGBERG, C. F. X., HEDBERG, R., & BAXTER, B. Management action based on one versus two dimensions of a job satisfaction questionnaire. *Personnel Psychology*, 1962, **15**, 145-150.

RECEIVED: June 21, 1972