

COMPARISON OF THE RARE RESPONSE AND VERTICAL PERCENT METHODS FOR SCORING THE BIOGRAPHICAL INFORMATION BLANK

MICHAEL G. AAMODT AND WALTER L. PIERCE, JR.
Radford University

This study investigated the relative effectiveness of two methods for scoring biographical information blanks. Data from five separate samples were weighted using the vertical percent method advocated by England (1971) and the rare response method suggested by Telenson, Alexander, and Barrett (1983). The results revealed that vertical percent scoring yielded significant validity coefficients for all five samples, while rare response scoring yielded significant validity coefficients for only two of the five samples. The results were discussed in terms of future applications for constructing valid biographical information blanks.

WHILE research has been abundant in demonstrating the effectiveness of the biographical information blank (BIB) in predicting various aspects of employee performance (e.g., Cascio, 1976; McClelland and Rhodes, 1969; Baehr and Williams, 1968), relatively little research has been conducted to determine the optimal methodology that should be used to construct and weight the BIB. Research that has investigated optimal BIB methodology has focused on such issues as rational scoring (Mitchell and Klimoski, 1982), preselection (Myers and Errett, 1959), and contingency items (Mahoney, 1958).

Most recently, Telenson, Alexander, and Barrett (1983) have suggested that rare response scoring, a method previously used

The authors would like to thank Kay Harrison for help in collecting the counselor data and Neil Hauenstein for his comments on earlier versions of this manuscript.

Requests for reprints should be addressed to Mike Aamodt, Department of Psychology, Radford University, Radford, VA 24142.

primarily in clinical work (Miner, 1965), is superior to the traditional vertical percent method of weighting BIBs (England, 1971). Basically, the rare response method involves assigning weight to responses on the basis of their deviation from the norm. Responses which are chosen by 15% or less of the sample are awarded two points, responses chosen by 15.1 to 30 percent of the sample are awarded one point, while responses which are chosen by more than 30 percent of the sample are not awarded any points. In contrast, vertical weighting assigns weights based on response differences between high and low performers. The greater the differentiation, the higher the weight.

In their study, Telenson et al. (1983) found that BIBs developed with vertical and horizontal weighting strategies and subjected to correlational analysis did not significantly predict any of three separate sales criteria. A BIB developed with rare response scoring resulted in a small but significant correlation ($r = .17$) in only one of the three sales criteria. Further Chi-square analysis indicated that the rare response method was able to significantly predict criterion group membership for all three criteria.

Based on their results, Telenson et al. (1983) concluded that due to validity and practicality considerations, the rare response method was a promising procedure for scoring BIB data. While this conclusion does not seem unreasonable, it should be noted that there are a number of questions regarding the Telenson et al. (1983) study which limit the conclusions that can be drawn from it.

First, it was surprising that the vertical weighting scheme did not produce a validity coefficient that was significantly different from zero. Previous research has clearly demonstrated that BIBs scored with the vertical percent method usually yield significant validity coefficients. Perhaps the reasons for the lack of significance can be found in the method used by Telenson et al. . The vertical weights were derived on a sample of only 38 sales people. This is far below the recommended derivation sample size of 150 suggested by England (1971) and Cascio (1982). The low sample size problem becomes magnified when one considers that Telenson et al. used the 38 subjects to weight 370 alternatives contained in the BIB.

Second, a hold-out sample was utilized for the vertical percent method but not for the rare response method. That is, 402 subjects were used to derive the rare response weights and 92 of those 402 were again used in the validation sample. However, 38 subjects were used to derive the vertical percent weights and these weights were validated on a *separate* sample of 92 subjects. Thus the rare response data were subject to not only the "fold-back" error

discussed by Cascio (1982) and warned against by England (1971), but also to the preselection effects discussed by Myers and Errett (1959).

Third, no direct comparison was made between the vertical percent method and the rare response method for either the validity coefficients or the hit-rate percentages. That is, Telenson et al. (1983) tested each method to determine whether or not it was significantly different from what would be expected by chance alone. However, the problem with using such an approach can be demonstrated in the following example.

Suppose that Method One resulted in a validity coefficient of .20 which was significant at the .05 level with the available number of subjects. Method Two resulted in a validity coefficient of .19 which was not significant at the .05 level. Without testing the difference between the two coefficients, it would not seem prudent to compare the two on the basis of one being significant and the other not. A test for differences between the two coefficients is needed before any claim for method superiority can be made.

With these problems in mind, it was the purpose of the present study to compare the effectiveness of BIB information developed by both the vertical weighting method suggested by England (1971) and the rare response method developed by Telenson et al. (1983). To do so, data from five separate samples were subjected to the two weighting procedures and the validity coefficients which resulted were compared.

Method

Subjects

Group 1: The subjects in this sample were 190 retail clerks at a large grocery store chain. BIB data was obtained through a 22-item questionnaire voluntarily completed by the employees. The criterion for this group was a composite score composed of supervisor ratings on two performance categories. Criterion groups were formed by placing employees with composite ratings of 7.0 or greater (out of 10.0) into the higher performance group and employees with composite ratings of less than 7.0 into the lower performance group.

Group 2: Subjects were 124 mental health counselors who were formally or currently employed at one of two mental health agencies. BIB data were obtained from employee files. Tenure was used as the criterion with employees with tenure of 18 months or greater

being placed into the higher tenure group and employees that quit or were terminated with less than 18 months tenure being placed into the lower tenure group.

Group 3: Subjects were 79 former workers at fast food restaurants. The BIB data were obtained through a 22-item questionnaire voluntarily completed by the subjects. The criterion was tenure, with subjects having tenure greater than seven months placed into the higher tenure group and subjects with tenure of seven months or less placed into the lower tenure group.

Group 4: Subjects were 62 ranchers who raised turkeys for a poultry processing plant. BIB data were obtained from company records. Number of turkeys raised served as the criterion and criterion groups were determined by splitting the sample into equal halves.

Group 5: Subjects were 32 industrial plant security officers. BIB data were obtained from a 22-item questionnaire voluntarily completed by the subjects. Supervisor ratings were utilized as the criterion with officers receiving ratings of excellent placed into the higher performance criterion group and officers receiving ratings of good placed into the lower performance criterion group (no officers received ratings of less than good).

Procedure

For each of the five samples, BIB data were weighted using the rare response guidelines mentioned in the Telenson et al. (1983) article and with the vertical percent guidelines listed by England (1971). For each of the two methods and in each of the five samples, item weights were derived, composite scores formed for each subject, and subject composite scores correlated with subject criterion group membership (e.g., Cascio, 1976; Schuh, 1967; McGrath, 1960). It should be pointed-out that in not using a hold-out sample to validate the two weighting methods, our data could be accused of suffering from the same "fold-back" error on which we questioned the Telenson et al. data. However, hold-out samples were not utilized for three reasons.

First, none of the subject groups had the minimum sample size of 250 suggested by Cascio (1982) and England (1971) for proper cross-validation of BIB data. Second, Schmitt, Coyle, and Rauschenberger (1977) have theorized that sampling error is smallest when samples are not split to form derivation and cross-validation or hold-out groups. Third, in the present study, any fold-back error that may or may not be present should affect the two weighting schemes about equally.

TABLE 1
Rare Response and Vertical Percent Validity Coefficients

Sample	Weighting Method		Z
	Rare Response	Vertical	
Grocery Clerks	-.07	.24**	1.70, <i>p</i> < .09
Counselors	.25**	.43**	1.28, <i>p</i> < .17
Fast Food Workers	-.05	.57**	3.03, <i>p</i> < .01
Turkey Raisers	-.13	.58**	2.97, <i>p</i> < .01
Industrial Police	-.42*	.76**	2.67, <i>p</i> < .01

* *p* < .05. ** *p* < .01.

Results and Discussion

As shown in Table 1, vertical percent scoring resulted in significant validity coefficients for all five samples. Rare response scoring resulted in significant coefficients in only two of the five samples. In no case did the rare response coefficient exceed that of the vertical percent method. Using Fisher's *Zr* transformation (Ferguson, 1981), the vertical percent coefficients were significantly different from the rare response coefficients for three of the samples.

The results from the present study seem to suggest that even though the rare response method resulted in two significant validity coefficients, the method cannot be considered superior to the vertical method as was suggested by Telenson et al. (1983). If any conclusion can be drawn, it would be that from a validity standpoint, the vertical percent method is superior to the rare response method.

A close examination of the results from this study suggest that the rare response method might actually be a measure of conventionality, rather than a method for scoring biodata. Rare response scoring resulted in a significant negative correlation with the criterion in a job (security police) which has long been noted for its conventional membership (Aamodt and Kimbrough, 1985) and resulted in a positive correlation with the criterion in a job (mental health counselor) which is characterized by a less conventional membership.

Another argument against the rare response method being used to weight data for all jobs involves the very theory upon which it is based. Telenson et al. (1983) argue that unusual information often is the most informative and from an impression formation aspect, they are probably correct (Myers, 1983). However, while rare responses may be informative, they do not necessarily imply better performance. Telenson et al. stated that "alternatives that logically implied a negative connotation with respect to sales performance were given a negative weight; otherwise, the weights were positive (pp. 75-76).

This type of procedure brings about obvious problems regarding what is considered a logical negative connotation. This is especially true considering that Mitchell and Klimoski (1982) have demonstrated the difficulty in using a rational approach in constructing a BIB. Perhaps an empirical assignment of positive or negative weights would be a solution.

In summary, the results of the present study suggest that unless further research demonstrates otherwise or offers methodological improvements, the vertical percentage method of weighting BIB items should be utilized. At present, problems with rare response scoring are too numerous to advocate its use for all jobs. Furthermore, unlike research investigating various schemes for weighting multiple predictors (Aamodt and Kimbrough, 1985; Wainer, 1976) and multiple criteria (Fralicx and Raju, 1982), the findings from this study strongly suggest that the method chosen to weight BIB data is important and will affect the outcome of the validation procedure.

REFERENCES

- Aamodt, M. G. and Kimbrough, W. W. (1985). Comparison of four methods for weighting multiple predictors. *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, 45, 477-482.
- Aamodt, M. G. and Kimbrough, W. W. (1985). Personality differences between police and fire applicants. *Journal of Police and Criminal Psychology*, 1, 10-13.
- Baehr, M. E. and Williams, G. B. (1968). Prediction of sales success from factorially determined dimensions of personal background data. *Journal of Applied Psychology*, 52, 98-103.
- Cascio, W. F. (1982). *Applied Psychology in Personnel Management*. Reston, VA: Reston Publishing Company.
- Cascio, W. F. (1976). Turnover, biographical data, and fair employment practice. *Journal of Applied Psychology*, 61, 576-580.
- England, G. W. (1971). *Development and use of weighted application blanks*. Minneapolis, MN: University of Minnesota Industrial Relations Center.
- Ferguson, G. A. (1981). *Statistical analysis in psychology and education*. New York: McGraw-Hill Company.
- Fralicx, R. D. and Raju, N. S. (1982). A comparison of five methods for combining multiple criteria into a single composite. *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, 42, 823-827.
- McClelland, J. N. and Rhodes, F. (1969). Prediction of job success for hospital aides and orderlies from MMPI scores and personal history data. *Journal of Applied Psychology*, 53, 49-54.
- McGrath, J. J. (1960). Improving credit evaluation with a weighted application blank. *Journal of Applied Psychology*, 44, 325-328.
- Mahoney, T. A. (1958). Weighted application blank analysis of contingency items. *Journal of Applied Psychology*, 42, 60-62.

- Miner, J. B. (1965). *Studies in management education*. New York: Springer.
- Mitchell, T. W. and Klimoski, R. J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology*, 67, 411-418.
- Myers, D. G. (1983). *Social Psychology*. N Y: McGraw-Hill.
- Myers, J. H. and Errett, W. (1959). The problem of preselection in weighted application blank studies. *Journal of Applied Psychology*, 43, 94-95.
- Schmitt, N., Coyle, B. W., and Rauschenberger, J. A. (1977). A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. *Psychological Bulletin*, 84, 751-758.
- Schuh, A. J. (1967). Application blank items and intelligence as predictors of turnover. *Personnel Psychology*, 20, 59-63.
- Telenson, P. A., Alexander, R. A., and Barrett, G. V. (1983). Scoring the biographical information blank: A comparison of three weighting techniques. *Applied Psychological Measurement*, 7, 73-80.
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213-217.