Scoring the Biographical Information Blank: A Comparison of Three Weighting Techniques

Paul A. Telenson, Ralph A. Alexander, and Gerald V. Barrett
The University of Akron

This study compared three techniques for scoring a biographical information blank (horizontal percent method, vertical percent method, and rare response weighting) against various criteria for field sales representatives. The comparisons were cross-validated over five consecutive time periods. The results showed that the rare weighting technique significantly predicted criterion group membership better than chance. Neither the horizontal nor vertical percent methods predicted criterion group membership better than chance. Based on predictive efficiency, the rare weighting technique was found to be superior to the other two techniques.

Since the work of Goldsmith (1922), the biographical information blank (BIB) has been successfully used to predict a wide variety of criteria. At a conference conducted solely for the discussion of biographical data, Henry (1966) concluded that "with very few exceptions it (BIB) has been found to be the best single predictor of future behavior where the predicted behavior is of a total or complex nature" (p. 248). Criteria investigated with BIBs have included such diverse areas as tenure or turnover, research competence and creativity, student achievement, and credit risk. BIBs have been developed for a wide variety of occupations, for example, sales, air crew personnel, seasonal em-

ployees, office personnel, scientists, factory workers, and managers. These categories are by no means exhaustive; they do, however, demonstrate the broad range of areas investigated with biographical information.

The BIB has many potential advantages over traditional prediction instruments. According to Guion (1965), there are two clear advantages. The first is that it is a relatively inexpensive instrument. After it has been developed and scored, continued use involves little expense to the organization. The second is that it is less likely to be distorted than other personality inventories. It is difficult for the applicant to fake his or her responses in such a manner as to make himself or herself "look good" (Kavanagh & York, 1972). Responses to verifiable items are surprisingly accurate (Cascio, 1975; Keating, Patterson, & Stone, 1950; Mosel & Cozan, 1952).

The BIB also augments traditional selection techniques. Owens and Henry (1965) concluded that the BIB is an extension of the application blank (which has already gained wide acceptance by organizations) and that it relates directly to traditional selection techniques. Dunnette, Kirchner, Erickson, and Banas (1960) suggested that a BIB used in conjunction with the traditional interview should improve the validity of the interview.

Other studies have shown that the BIB can satisfy the legal requirements of a valid selection

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 7, No. 1, Winter 1983, pp. 73-80
© Copyright 1983 Applied Psychological Measurement Inc.
0146-6216/83/010073-08\$1.40

instrument. Moore (1968) demonstrated that the BIB used in his study was a powerful selection instrument which was not discriminatory. Cascio (1976) found that turnover can be accurately predicted by biographical information within the requirements of fair employment practices.

Despite its popularity, the use of biographical information has not been without criticism. Dornon (1970) reported that past BIB investigations have generally employed poor methodology. He reported that such studies tended to include insufficient detail, small sample sizes, the use of questionable statistics, and poor validation procedures. Others have reported shrinkage in validity over time (Dunnette et al., 1960; Wernimont, 1962). Little concrete evidence has been presented that demonstrates the stability of predictions based on BIB scores.

Another problem is the relative lack of comparative research on weighting techniques. Even though many unique weighting procedures have been proposed, there have been few instances of studies where different scoring systems were compared against each other to determine if one system was more predictive than another. Weighting techniques range from sophisticated statistical procedures possible only through the use of computers to simpler techniques requiring minimal calculations.

One of the more publicized weighting procedures was developed by England (1971) and can be traced back to the "vertical percent" method of Stead and Shartle (1940). This technique was based on tables developed by Strong (1926). The tables, originally developed for use with the Strong Vocational Interest Blank, are used to determine the weights for items as a function of the percentages of individuals from different criterion groups responding to the item.

A similar weighting system known as the "horizontal percent" method, which requires less calculation than the vertical percent method, was also proposed by Stead and Shartle (1940). Although no studies were identified that used the horizontal weighting technique, it may be a useful technique because of its ease of scoring.

A third method, rare response scoring, is quite common in clinical diagnostic instruments and projective tests although not widely discussed in the psychometric literature. Two quite different rationales are proposed for the use of rare response keying. Berg (1967) has suggested that deviant responses, irrespective of item content, represent a pervasive response set of the nonconforming personality. Another approach has been that unusual or unexpected responses often have greater informational value with respect to an individual's standing on a particular characteristic. It is this latter argument that is put forth for the two projective measures designed to be used for differentiating within the normal personality range (e.g., for employee selection): the Miner Sentence Completion Scale (Miner, 1965) and the Picture Arrangement Test (Tompkins & Miner, 1957). Like the horizontal weighting technique, no evidence of using the rare weighting technique in BIB research was located in the literature.

A number of other statistical weighting schemes have been used for developing BIB weights. Such criterion-related systems as multiple regression, and correlation, or phi coefficient, weighting systems were not used in the present study because of the small sample size available for developing weights. The vertical and horizontal percent methods used here do not appear to be as susceptible to sampling fluctuation of the derived weights (England, 1971). The purpose of this study, then, was to compare vertical percent, horizontal percent, and rare response scoring techniques for a BIB against performance criteria for field sales representatives.

Method

Sample

The sample consisted of 402 male applicants to a large company manufacturing consumer goods. Of these, 130 field sales personnel were eventually hired. Only males were included in this study due to the small number of females.

The total sample of 402 ranged in age from 21 to 59, with an average age of 30. The average attained education level was 16 years.

Criteria

From the applicants who were hired as sales personnel, sales performance criteria were collected consisting of sales, sales-quota ratio, and incentive payments. A terminal rating criteria based on promotions, demotions, terminations, and transfers was available for those individuals who left sales positions after a minimum of one full semi-annual period. Criterion data were obtained for six consecutive semi-annual periods.

Criterion groups (high, medium, and low) were determined by dividing the sales force into thirds separately for each of the three sales performance criteria for each time period. For each individual who left a sales position, a terminal rating was established. For those who were promoted or who quit and were considered a loss to the organization, the terminal rating was "high." Those who were transferred or who quit with a final evaluation of "average" were rated "medium," and those who were fired or demoted or who quit with a low evaluation were rated "low."

Biographical Information Blank

All applicants were required by the company to complete a 69-item multiple-choice BIB prior to being considered for employment. Each alternative to each question was analyzed, resulting in 370 alternatives.

Scoring keys for the BIB were developed by the vertical and horizontal percent methods separately for each of the three sales performance criteria (sales, sales-quota ratio, and incentive payments) using the sales sample ($N = 38$) available during the first semi-annual period of the study. For each criterion measure, persons in the top one-third were identified as the high criterion group and those in the bottom one-third as the low criterion group.

Scoring keys for the vertical percent method were constructed following England's (1971) procedure. For each alternative in the BIB, the percentage of each of the two criterion groups responding to that alternative was found and the difference between those two percentages calculated. A net weight was assigned to that difference based on Strong's (1926) tables; if the percentage was greater for the low criterion group, the weight was given a negative value. These weights were then transformed by assigning a value of 0 to weights of -4 or lower, a value of 1 to weights of -3 to $+3$, and a value of 2 to weights of $+4$ or greater. England (1971) recommends this final step of transforming the weights to reduce the impact of sampling fluctuation in percentages. Finally, any question on the BIB for which all alternatives received the same final value was eliminated from scoring.

For the horizontal percent method, three scoring keys, based on high and low criterion groups, were constructed following Stead and Shartle's (1940) procedure. For each alternative of the BIB, a percentage was calculated for the high criterion group by dividing the number of high criterion group responses by the total number of persons responding in both the high and low criterion groups. A final weight for that alternative was then calculated by multiplying the high criterion group percentage by 10. Questions for which all alternatives received the same weight were eliminated from scoring.

The rare scoring technique was a modification of the method described by Tompkins and Miner (1957). Weights were assigned to items based on the frequency with which they were responded to in the total applicant sample. A weight of 0 was assigned to an item if it was responded to by 30% or more of the applicants. A weight of 1 was assigned to items that were responded to by 15% to 30% of the applicants, and a weight of 2 was assigned to those items that were responded to by less than 15% of the applicants. Alternatives that logically implied a negative connotation with respect to sales performance were given a negative weight; otherwise,

the weights were positive. Questions that had all alternatives receiving the same assigned weight were excluded from the analysis.

Cutoff scores for each of the seven scoring keys were determined by dividing the relevant weighting sample into thirds (high, medium, and low) based on total scores for each scoring key.

It might seem that the most straightforward evaluation of the predictive usefulness of each of these keying techniques would be the correlation of BIB total scores against each of the raw criterion values. Although such an analysis was performed, the results are of limited value. The sales performance criteria are somewhat contaminated by (1) each field sales representative being responsible for a fixed geographic area and (2) at least some of the performance differences among sales personnel being a function of the assigned area. In addition, changing economic conditions, advertising efforts by the company, and other factors outside the salespersons' control over the time span of the study influenced the raw criterion values to some extent. Therefore, the principal analysis of the predictive efficiency of each scoring system involved predicting criterion group membership from BIB scores. For each analysis, a 3×3 table was developed based on predicted group membership (high, medium, and low BIB scores) and actual group membership (high, medium, and low criterion scores). Table 1 gives a representation of a completed table.

Several of these cells warrant further discussion. In the eyes of the company, an individual who is predicted to be a high performer but turns out to be a medium performer would not be considered a true loss to the organization. Thus, Cell 4 was considered a "hit." Likewise, if an individual is predicted to be a medium performer in the company and turns out to be a high performer, the organization faces no loss. This was the rationale for labeling Cell 2 a "hit." On the other hand, a predicted low performer who turns out to be a medium or high performer would result in a loss to the organization; a good performer would not have entered the company.

Thus, Cells 3 and 6 were labeled "miss." For statistical analysis, therefore, this 3×3 table was collapsed into a 2×2 cross-classification of successful versus unsuccessful.

Results

Of the initial sample of 130 field sales representatives, sales performance criterion data were available for the first semi-annual period on 38 of them and that group was used for the criterion-related keying sample. The validation analyses that follow involved only the remaining sample of 92. The correlations among the raw BIB scores and the performance criteria for the first available semi-annual period for each sales person are presented in Table 2.

As can be seen from the table, the correlation results are not particularly strong. Only one of nine coefficients (rare score versus sales; $r = .17$, $p < .05$) is statistically significant. The rare score key does have the highest validity coefficients against each of the three criteria. These weak correlation results were expected because of the contaminants discussed earlier.

The data from the correlational analysis was submitted to chi-square analyses of the 2×2 cross-classifications of predicted versus actual criterion group membership of successful versus unsuccessful sales performance. Those results are summarized in Table 3, which shows the actual and expected (chance) hit rates and percentages of correct classification. As can be seen from those results, the chi-square values for the three rare score analyses are statistically significant while none of the others are.

Of the 92 salespersons included in the validation sample, 54 left sales positions at some time during the study. The terminal ratings of these persons was used as the criterion for similar chi-square analyses. Neither the vertical nor horizontal percent method was statistically significant. In fact, for both, the actual hit rate was less than the expected (chance) hit rate. For the rare score method, the hit rate of 63% correct classifications against chance expectancy of 52%

Table 1
Hit/Miss Classifications

Actual (Criterion Scores)	Predicted (Biographical Information) (Blank Score)		
	High	Medium	Low
High	Hit (1)	Hit (2)	Miss (3)
Medium	Hit (4)	Hit (5)	Miss (6)
Low	Miss (7)	Miss (8)	Hit (9)

was significant [$\chi^2 (df = 1) = 4.162; p < .05$].

Finally, actual hit rate percentages were compared with the expected hit rate percentages using each scoring technique for all available criterion data across consecutive time periods. Those results are presented in Table 4. The Wilcoxin sign rank test (Wilcoxin, 1949) was used to determine if the actual hit rate percentages significantly differed from the expected hit rate percentages using each scoring system.

Table 4 reveals that the biographical information blank scored by the vertical percent method did not consistently predict criterion group membership better than chance. The actual hit rate percentages were not consistently higher than the expected hit rate percentages. When actual hit rate percentages were compared with expected hit rate percentages using the Wilcoxin sign rank test, no significant differences were found.

Table 4 also shows that the biographical information blank scored according to the horizontal

percent technique did not successfully predict criterion group membership better than chance. No significant differences were found between actual hit rate percentages and expected hit rate percentages based on Wilcoxin's sign rank test.

Finally, as can be observed in Table 4, the biographical information blank scored by the rare weighting technique consistently predicted criterion group membership better than chance. In all cases, the actual hit rate percentages for the rare weighting technique were higher than the expected hit rate percentages. Using Wilcoxin's sign rank test, a significant difference ($p < .01$) was obtained between actual and expected hit rate percentages.

Discussion

The results of this study demonstrate that the rare weighting technique consistently predicted criterion group membership better than chance and that the horizontal percent method and

Table 2
Correlations of Raw BIB Scores with Raw Sales Performance
Criteria for the First Available Semi-Annual Period
for Each Salesperson (N = 92)

Sales Performance Criterion	Scoring Key Method		
	Vertical	Horizontal	Rare
Sales	-.07	-.05	.17*
Sales-quota Ratio	.04	.07	.13
Incentive Payment	.06	.04	.15

* $p < .05$

Table 3
Hit-Rate Analysis of Cross-Classification Predicted (P) vs.
Actual (A) of Criterion Group Membership (N = 92)

Criterion Group	Hit-Rate				χ^2
	N		%		
	P	A	P	A	
Vertical Percent Method					
Sales	55	55	60	60	.006
Sales-Quota Ratio	61	61	66	66	.004
Incentive Payment	63	61	68	66	2.051
Horizontal Percent Method					
Sales	55	61	60	66	1.611
Sales-Quota Ratio	59	58	64	63	.063
Incentive Payment	61	58	66	63	.987
Rare-Score Method					
Sales	65	56	71	61	6.392*
Sales-Quota Ratio	64	55	70	60	6.501*
Incentive Payment	62	55	67	60	4.289*

* $p < .05$ (df = 1)

England's vertical percent technique did not predict criterion group membership better than chance.

There are several implications for these findings. First, this study presented one of the few attempts where different biographical information blank weighting techniques were compared. In earlier studies that did compare weighting techniques (McGrath, 1960; Scollay, 1956), only one criterion was investigated and only one cross-validation sample was evaluated. Therefore, actual comparisons were somewhat limited. In this study, three weighting systems were compared based on several criteria obtained from five time periods. The resulting comparisons were, therefore, more thorough than earlier attempts at comparing weighting systems.

In addition, the rare weighting system consistently predicted criterion group membership across consecutive time periods. Earlier studies have reported a drop in the predictive efficiency of the biographical information blank over time (Dunnette et al., 1960; Wernimont, 1962). No such drop in predictive efficiency was found in this study using the rare weighting technique.

Rare weighting, as the result of its scoring ease, is also of practical significance. Even though, by its nature, the rare weighting technique was easier to implement than the other two techniques, it was more efficient in predicting criterion group membership. Lawshe and Schucker (1959) have presented similar evidence regarding simple weighting techniques. The results of this study indicate that the organization would have averaged 62% correct classifications of sales personnel into high, medium, and low sales criterion groups.

Finally, an interpretation of the items that were included in the rare weighting technique logically indicated good and poor performers. From a construct validity point of view, items indicating negative connotation were weighted negatively in the rare weighting technique. No such hypotheses can be made with criterion-related techniques such as the other two compared in this study. This further vindicates Nunnally's (1978) admonition against criterion-related test construction and scoring.

It must be recognized that the sample available for deriving the criterion-referenced keys

Table 4
Hit-Rate Percentages for All Three Methods

Method, Criterion, and Type	Time Period				
	1973		1974		1975
	First Half (N=28)	Second Half (N=35)	First Half (N=53)	Second Half (N=76)	First Half (N=90)
Vertical Percent Method					
Sales					
Actual	43%	40%	51%	51%	48%
Expected	45%	47%	52%	53%	55%
Incentive Payment					
Actual	39%	43%	51%	54%	58%
Expected	43%	46%	53%	55%	59%
Sales/Quota Ratio					
Actual	46%	51%	58%	57%	60%
Expected	43%	46%	52%	56%	58%
Horizontal Percent Method					
Sales					
Actual	39%	37%	45%	50%	52%
Expected	45%	47%	52%	53%	55%
Incentive Payment					
Actual	50%	51%	57%	58%	57%
Expected	44%	47%	52%	54%	56%
Sales/Quota Ratio					
Actual	43%	43%	53%	51%	51%
Expected	45%	47%	52%	54%	57%
Rare Response Method					
Sales					
Actual	57%	54%	64%	65%	57%
Expected	45%	47%	52%	53%	55%
Incentive Payment					
Actual	57%	54%	60%	62%	60%
Expected	45%	47%	52%	53%	55%
Sales/Quota Ratio					
Actual	57%	54%	57%	57%	58%
Expected	45%	47%	52%	53%	55%

was not large enough to produce stable weights and that this is one likely reason for their failure to validate. These results, however, point up a practical advantage of rare score over criterion-related keying methods. In actual field settings where the BIB is likely to be investigated for use in selection systems, it would be unusual to find

a sample of present employees of sufficient size to develop a stable criterion-related key. Much more common is the situation faced in this study—a relatively small validation sample but a substantial applicant pool for developing a stable set of rare-score weights. It is encouraging, therefore, to find that a construct valid rare-

score key does hold up reasonably well on validation. It would be interesting to see whether or not such a rare-score technique could be substantially improved by further optimizing the weighting scheme and by the explicit inclusion of alternatives in the BIB that would conform to the logic of rare-score weighting. That is, the specific inclusion of items that would be considered normal behavior (positive or negative with respect to the job performance of interest) but non-normative in the sense that few people would be expected to select the item.

References

- Berg, I. A. *Response set in personality assessment*. Chicago: Aldine, 1967.
- Cascio, W. F. Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, 1975, 60, 767-769.
- Cascio, W. F. Turnover, biographical data, and fair employment practice. *Journal of Applied Psychology*, 1976, 61, 576-580.
- Dornon, J. M. Identification of long-tenure hourly factory workers using a weighted application blank. *Experimental Publication System*, 1970, 8. (Ms. No. 276-2)
- Dunnette, M. D., Kirchner, W. K., Erickson, J. R., & Banas, P. A. Predicting turnover of female office employees. *Personnel Administration*, 1960, 23, 45-50.
- England, G. W. *Development and use of weighted application blanks* (rev. ed.). Dubuque IA: William C. Brown, 1971.
- Goldsmith, D. B. The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology*, 1922, 6, 149-155.
- Guion, R. M. *Personnel testing*. New York: McGraw-Hill, 1965.
- Henry, E. R. Conference on the use of biographical data in psychology. *American Psychologist*, 1966, 21, 247-249.
- Kavanagh, M. J., & York, D. R. Biographical correlates of middle managers' performance. *Personnel Psychology*, 1972, 25, 319-332.
- Keating, E., Patterson, D. G., & Stone, C. H. Validity of work histories obtained by interview. *Journal of Applied Psychology*, 1950, 34, 6-11.
- Lawshe, C. H., & Schucker, R. E. The relative efficiency of four test weighting methods in multiple prediction. *Educational and Psychological Measurement*, 1959, 19, 103-114.
- McGrath, J. J. Improving credit evaluation with a weighted application blank. *Journal of Applied Psychology*, 1960, 44, 325-328.
- Miner, J. B. *Studies in management education*. New York: Springer, 1965.
- Moore, C. L. *Ethnic differences as measured by a biographical inventory questionnaire*. Greensboro NC: The Creativity Research Institute of the Richardson Foundation, 1968.
- Mosel, J. N., & Cozan, L. W. The accuracy of application blank work histories. *Journal of Applied Psychology*, 1952, 36, 365-369.
- Nunnally, J. M. *Psychometric theory* (2nd ed.). New York: McGraw-Hill, 1978.
- Owens, W. A., & Henry, E. R. *Biographical data in industrial psychology: A review and evaluation*. Greensboro NC: The Creativity Research Institute of the Richardson Foundation, 1965.
- Scollay, R. W. Validation of personal history items against a salary increase criterion. *Personnel Psychology*, 1956, 9, 325-335.
- Stead, N. H., & Shartle, C. L. *Occupational counseling techniques*. New York: American Book Company, 1940.
- Strong, E. K. An interest test for personnel managers. *Journal of Personnel Research*, 1926, 5, 194-203.
- Tompkins, S. S., & Miner, J. B. *The Tompkins-Horn Picture Arrangement Test*. New York: Springer, 1957.
- Wernimont, P. F. Re-evaluation of a weighted application blank for office personnel. *Journal of Applied Psychology*, 1962, 46, 417-419.
- Wilcoxon, F. *Some rapid approximate statistical procedures*. Stanford CT: American Cyanamid Company, 1949.

Acknowledgments

This research is based on the master's thesis of the first author, who is currently a Training Specialist with Tremco, Inc., 10701 Shaker Boulevard, Cleveland, Ohio.

Author's Address

Send requests for reprints or further information to Ralph A. Alexander, Department of Psychology, The University of Akron, Akron OH 44325.