Weighted Application Blanks: An Empirical Approach for the Selection of Delivery Personnel Psychology

Barton Lee Dahmer

Follow this and additional works at: https://digitalcommons.wku.edu/theses

Part of the Human Resources Management Commons, and the Industrial and Organizational Psychology Commons

Recommended Citation

https://digitalcommons.wku.edu/theses/2082

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
Dahmer,

Barton Lee

1985
WEIGHTED APPLICATION BLANKS:
AN EMPIRICAL APPROACH FOR THE SELECTION
OF DELIVERY PERSONNEL

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

By
Barton Lee Dahmer
September, 1985
AUTHORIZATION FOR USE OF THESIS

Permission is hereby granted to the Western Kentucky University Library to make, or allow to be made photocopies, microfilm or other copies of this thesis for appropriate research or scholarly purposes.

☐ reserved to the author for the making of any copies of this thesis except for brief sections for research or scholarly purposes.

Signed B.L. Palmer

Date 10-25-85

Please place an "X" in the appropriate box.

This form will be filed with the original of the thesis and will control future use of the thesis.
WEIGHTED APPLICATION BLANKS:

AN EMPIRICAL APPROACH FOR THE SELECTION

OF DELIVERY PERSONNEL

Recommended November 8, 1985
(Date)

Elizabeth S. Effmeyer
Director of Thesis

David L. Rossen
Ray M. Mendel

Approved December 4, 1985
(Date)

Elmer Gray
Dean of Graduate College
Acknowledgements

Acknowledgement is made for the assistance and guidance received from my thesis committee: Betsy Erffmeyer, Chairperson, Ray Mendel, and Dan Roenker.

Special appreciation is expressed to my brother, Steve Dahmer, whose insight and assistance with this thesis was invaluable. Sincere appreciation is also expressed to the Personnel Director and Staff of IDC for affording me the opportunity to access the data which made this thesis possible.

Special thanks are offered to Bob Cobb and Dennis Turner for their assistance with data input and computer analyses.

Finally, my deep appreciation is expressed to Sharon Dahmer, my wife, whose encouragement helped me to persist during the many long months of work.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................ iii
ABSTRACT ................................................................ v

CHAPTER I
  Introduction and Literature Review .................. 1

CHAPTER II
  The Present Study ......................................... 18

CHAPTER III
  Methodology ............................................. 22

CHAPTER IV
  Results ..................................................... 33

APPENDIX A ................................................. 40
APPENDIX B ................................................. 46
APPENDIX C ................................................ 50
APPENDIX D ................................................ 53
REFERENCES ............................................... 56
WEIGHTED APPLICATION BLANKS:
AN EMPIRICAL APPROACH FOR THE SELECTION
OF DELIVERY PERSONNEL

Barton Lee Dahmer September, 1985 60 pages
Directed by: E. Erffmeyer, R. Mendel, and D. Roenker
Department of Psychology Western Kentucky University

The purpose of this study was to develop and validate a biographically weighted application blank for use in selecting delivery personnel. The England (1971) procedure for weighting biographical information was utilized. The criterion was a ratio of planned work time to actual work time. It was hypothesized that (a) significant derivation and cross-validities would be obtained, and (b) significant practical benefits in terms of correct placement of workers in the high and low criterion groups would result. The first hypothesis was partially supported in that a significant derivation validity was obtained ($r = -.56$, $P < .05$). However, the cross-validity ($r = -.12$) was not significant. Thus, the second hypothesis was not supported.

A discussion of the results and recommendations for the implementation of the weighted application blank are provided.
CHAPTER I
INTRODUCTION AND LITERATURE REVIEW

If they gave an award for the 'most consistently valid predictor,' biographical information would be the winner. Of all the predictors used to forecast job performance, biographical information (as a general class) has consistently shown the greatest validity. What is more remarkable about this finding is that it occurs across wide differences in people, jobs, and criteria. (Muchinsky, 1983 p. 124.)

Even a cursory review of the literature supports Muchinsky's statement of the utility of biodata as a predictor (Mumford, Cooper, & Schemner, 1984). Asher and Sciarrino (1974) noted that biographical data has a higher percentage of validity coefficients equal to or above the .50 level than do work samples, intelligence tests, mechanical aptitude tests, finger dexterity tests, personality tests, or spatial relations tests. A review of 38 biographical studies by England in 1971 revealed that 95% of the studies yielded significant cross-validity coefficients. The validity estimates ranged from .19 to .83 and the median was .45. Asher
(1972) conducted a similar review of biographical data in which 31 cross-validity estimates were examined. Of those, 97% of the validity estimates were .30 or higher, 74% were .40 or higher, 55% were .50 or higher, and 35% were .60 or higher (Cascio, 1982). Other examples of biographical data yielding high and significant cross-validity coefficients abound (Beatty & Schneier, 1981; Cascio, 1976, 1982; Owens, 1976; and Owens & Schoenfeldt, 1979). The cross-validity estimates derived in these studies were obtained across a diverse group of occupational areas and criteria such as college selection, job performance, productivity, absenteeism, rate of salary increase, number of publications, number of patents, creativity ratings, sales volume, credit risk, employee theft, and, most commonly, turnover.

When biographical data is used as a predictor "primary emphasis is placed on past behavior to predict future behavior." (Cascio, 1982, p. 194.) An example will help illustrate the point. Let us suppose that a company has a problem with employee performance, that is, a significant number of employees are not performing up to requirements. One way to help correct the problem would be to select applicants who are likely to perform well on the job. In order to use biodata as a predictor of performance, the actual criterion measure (performance) is first developed. The employees are
then divided into two groups based on the criterion measure: those that performed well on the job and those that performed poorly on the job. The next step is to identify how the groups differ in relation to past performance as indicted by the biographical data. For example, if the responses to some biographical questions are much more frequent in one of the two groups, those questions may be predictive of job performance. As an example, assume that 90% of the high performers graduated from college and only 10% of the low group graduated from college. In this case, the biographical question "Did you graduate from college?" would likely be a good predictor of performance. Weights would be assigned to the biographical options in direct proportion to the group differentiation identified. Each applicant would then be given a score by summing the appropriate weights assigned to each option chosen by the applicant. In this fashion, hiring decisions could be based on the total biographical scores obtained by each applicant. That is, generally the higher the total score the more likely the applicant is to perform well on the job. Of course this example is oversimplified, but the concept is illustrative of one method by which biographical data may be used to predict a criterion, in this case job performance. Each of the aforementioned steps as well as many others will be
examined in detail in the following pages.

There are currently two widely accepted approaches for systematically using biographical data as a predictor: the empirical approach (EA) and the rational approach (RA). Each approach has its strengths and its limitations. These attributes and the methodological procedures for each approach are examined in the next two sections of this paper. The review will serve as an explanation for the decision to use the EA for the development of a weighted application blank for an international delivery company.

The Empirical Approach

The empirical approach was the first systematic use of biographical data as a predictor. The first documented, large scale use of biographical data as a predictor was proposed in 1894 by Colonel Thomas L. Peters of the Washington Life Insurance Company (Owens, 1983). In order to improve the selection of insurance agents, Peters suggested that all applicants complete a standard list of biographical questions. The subsequent works of Goldsmith (1922) and Russell and Cope (1925) were the first published efforts of scored personal history data used for purposes of improving selection (Mitchell, & Klimoski, 1982). These same publications led to the development of the empirically keyed weighted application blank (WAB) which focuses on criterion-item
covariation (Mitchell & Klimoski, 1982). By 1935 the empirical approach (EA) was commonly used. Over twenty-three EA methods are currently available, with the WAB approach of England (1971) being the most commonly cited (Pace & Schoenfeldt, 1977; Weiss, 1976; Mitchell & Klimoski, 1982). As has already been stated, the EA has consistently yielded high derivation and cross-validities across diverse fields and criteria. In addition, the EA requires a relatively small amount of time to construct a weighted application blank. England (1971) estimated the construction time to be about 100 man-hours. However, Mitchell and Klimoski pointed out that with the increasing accessibility to computers, the 100 man-hour estimate is probably now high.

Two criticisms of the EA are particularly relevant. Schwab and Oliver (1974) suggested that the validity of the EA is substantially overstated in the literature. That is, many of the earlier studies did not perform cross-validation procedures. However, many of the more recent studies used cross-validation as a matter of course and still yielded high cross-validities (Mitchell, & Klimoski, 1982). Studies not employing cross-validation procedures may have spuriously high validity coefficients and should be viewed with caution (Schwab and Oliver).

A criticism which is often cited against the EA is that it lacks causal implications (Owens, 1983). The theoretical

The Rational Model

In response to the criticisms of the empirical approach, particularly the lack of theoretical interpretation, the rational approach (RA) for scoring the biodata questionnaire began to develop (Mitchell & Klimoski, 1982). Levine and Zachert (1951) are credited with being the first to explore the use of a more rationally based strategy for scoring biodata (Mitchell & Klimoski, 1982).

In its present form, the RA utilizes both theoretical considerations and internal consistency analyses to construct item composites from a pool of questionnaire items. The RA identifies factors or groups of items that measure an interpretable set of constructs. Prediction of criteria is accomplished by regressing the factors on the criteria (Hornick, James, & Jones, 1977).

Owens (1971) has modified the RA to include cluster analysis. By summing the scores of each individual within each factor developed during the factor analysis, a biodata profile for each individual is created. Cluster analysis (e.g., the minimum variance method by Ward and Hook, 1963) is then used to group or cluster similar profiles. Using quantifiable criteria such as mathematical or verbal ability, the performance level of each group is identified.
Individuals who are found to have biographical profiles similar to a specific group or cluster are then predicted to perform on the criteria as did the members of that group. Thus, the rational model as modified by Owens has the potential for attributing to the individual several behavioral characteristics of his assigned subgroup (Owens, 1971).

Proponents of the RA contend that theoretical benefits, that is, understanding what constructs the items are measuring, as well as practical benefits, that is, placement or selection, are gained from this method (Pace & Schoenfeldt, 1977). Indeed, in comparing 40 different studies using the RA, Owens and Schoenfeldt (1979) found that three fourths of them yielded significant positive validities.

Despite its potential for identifying theoretically meaningful constructs, the rational method has received some criticism. Some of the disadvantages of the RA are simple and undisputed. When the RA is employed, a relatively high level of knowledge and proficiency on the part of the developer as well as a healthy budget are required (Mitchell & Klimoski, 1982). In addition, an extremely large sample size is required (Owens, 1979). From a practical standpoint, it may not be necessary that every application of personnel technology contribute to its theoretical counterpart, especially if this leads to higher costs than
would be expected with equally effective alternatives (Mitchell & Klimoski, 1982).

Other more disputed criticisms of the psychometric properties of the RA have been set forth by several researchers. Bechtoldt (1959) and Cronbach and Meehl (1955) stated that construct validity, which is based only on inter-item and inter-test relationships, is circular. Although high inter-item and inter-test correlations indicate that the correlated items or tests are measuring the same construct, it is still not known whether the items or tests are measuring the construct that they are intended to measure. Thus, the theoretical implications derived from the results may not be applicable to the construct in question.

Another line of criticism of the RA focuses on its potential loss of power for prediction. Hornick et al. (1977) noted that the RA utilizes the commonality of items. This process suppresses unique item variance which would possibly overlap with the criterion and thus improve prediction. For example, consider an item which is independent of items in each of the derived factors yet still accounts for variance in the predictor. With the RA, an item such as this would probably not be included in the biographical questions because of the factor-analytic procedures.
Rational vs. Empirical Scoring

A review of the empirical and rational methods for scoring biodata was conducted to determine which method was better for use in the present study. The following section reviews three recent comparisons of the RA and the EA for keying biographical data. The results of three studies were examined to substantiate the decision to use the EA in the present study.

Hornick, James, and Jones (1977) compared the RA and the EA for scoring the Psychological Climate Questionnaire in a study using 398 male firemen as the subjects. The mean rating across ten areas of job performance as measured on the standard civil service performance evaluation was used as the criterion. However, there were actually two criterion measures: a rating completed nine months before the study and a rating completed three months before the study. The RA resulted in cross validities for the nine month and three month criteria of .46 and .42, respectively. Similarly, the EA resulted in cross validities for the two criteria of .47 and .44. There was no significant difference between the cross-validity coefficients of the EA and RA. Hornick et al. then randomly reassigned subjects within both the high and low criterion groups and repeated the process. The results were similar to the first outcome in that no statistically significant difference was found between the cross validities of the two approaches. Hornick
et al. also performed content analyses on the items used in each approach and found that only the RA offered a sound understanding of the item criterion relationships.

A similar study was undertaken by Mitchell and Klimoski (1982), who compared the cross-validities and the predictive accuracies of the RA and the EA in a real estate sales context. The attrition rate of the entry level sales personnel in real estate is over seventy-five percent. Hypothesizing that a properly validated biographical inventory could help solve the attrition problem, Mitchell and Klimoski compared the relative effectiveness of the RA and the EA in predicting successful candidates.

The criterion used by Mitchell and Klimoski (1982) was obtaining an Ohio license to sell real estate. The subjects were 698 students in prelicensing classes from five colleges and technical schools in Ohio. The procedure used in the two approaches was essentially identical to those outlined earlier in the paper. In the EA, 88 biodata items were created and subsequently answered by the subjects. During the weighting procedure the net weights were retained rather than converted to assigned unit weights. England (1971) noted that this process is perfectly acceptable. After the items were weighted, the WAB was cross validated and a cut off score was set. In the RA, 86 of the original 88 biodata questions were retained. Subsequent factor analysis resulted in six relatively uncorrelated dimensions. As the
final step in the RA, simultaneous multiple regression was used to derive a regression equation relating the component scores to the criterion.

The results of the study were quite decisive (Mitchell & Klimoski, 1982). The derivation validity was .355 for the RA and .592 for the EA. Moreover, the cross validities for the RA and EA were .362 and .462, respectively. The derivation and cross-validity estimates were compared via t-tests for dependent coefficients. Both the derivation and cross-validity estimates obtained with the EA were significantly greater than the derivation and cross-validity estimates obtained with the RA. In addition, the EA cross-validity represented an 8.2% improvement over the cross-validity of the RA. However, the shrinkage between the derivation validity and the cross-validity for the EA was found to be significant (p < .05), while shrinkage in the RA was virtually nonexistent.

In an attempt to assess the practical difference between the RA and the EA, Mitchell and Klimoski (1982) compared the results of using the cut off score of the EA and the regression equation of the RA in predicting the criterion group for the hold-out sample. It was found that the EA eliminated 59.4% of the failures and retained 80.5% of the successful personnel, while the RA eliminated 62.4% of the failures and retained 68.8% of the successful personnel. Thus, Mitchell and Klimoski noted that the EA
represented a 5.6% improvement in correct placement of employees.

A more recent study by Moomaw, Neiner, and Kruse (1984) compared the RA and the EA approach for predicting two turnover criteria, termination and rehire status, in entry and lower level technical/professional employees in 104 life insurance companies. Moomaw et al. found that the EA led to significantly higher derivation and cross validities than did the RA. The derivation validities for termination and rehire status obtained using the EA were .11 and .37 ($p < .001$), respectively. The cross validities obtained using the EA were .14 and .18 ($p < .001$), respectively. The derivation validities obtained using the RA were .09 ($p < .001$) for termination and .27 ($p < .001$) for rehire status. The cross validities obtained using the RA were .07 ($p < .03$) for termination and .08 ($p < .03$) for rehire status. The shrinkage which occurred with both methods was found to be significant at the .05 level. Moomaw et al. noted that from a purely statistical standpoint it appeared that the EA was the superior approach.

Both the Moomaw et al. (1984) study and the Mitchell and Klimoski (1982) study strongly suggest that the EA has both statistical and practical advantages over the RA if prediction is the major concern. Other factors such as time, expense, and technical skill undoubtedly favor the EA. On the other hand, the Hornick et al. (1977) study suggests
that the RA has advantages in the area of theoretical interpretation that may potentially have significant monetary and practical benefits. As Guion (1965) stated, an understanding of why a predictor is valid will enable researchers to improve it, to determine its limitations, and to safeguard against the danger of unsuspected or irrelevant discrimination. Even though the RA has much future potential, presently the approach appears to offer few immediate practical advantages over the EA. Thus, if prediction is the primary objective of a particular study, it would seem appropriate to use the less expensive, less time consuming, and less complex EA. However, all research is not undertaken solely for the purpose of prediction. Opportunities may arise where researchers have the technical skill, equipment, and the theoretical interest to employ the RA. In such cases it would seem advantageous to use the RA.

**Utility**

Organizations are frequently concerned with the utility of selection procedures. Executives often demand estimates of the costs and benefits associated with personnel programs such as selection procedures (Cascio, 1982). Until recently few selection devices were evaluated in terms of economic utility (Cascio). Utility is defined as the degree to which a selection device improves the quality of people selected beyond what would
have occurred had the selection device not been used (Blum & Naylor, 1968). In 1978, Hunter and Schmidt conducted a review of the economic utility of selection procedures and concluded that personnel psychologists often fail to realize the economic value of a valid selection device. Schmidt, Hunter, McKenzie and Muldrow (1979) noted that with the advent of new methods for estimating the standard deviation of the criterion in dollar terms, questions concerning the economic and productivity implications of valid selection procedures have recently come to the forefront in Industrial-Organizational Psychology.

In 1949 Brogden made a major advance in utility interpretations by using the principles of linear regression to demonstrate how the standard deviation of job performance in dollars, the validity coefficient, the selection ratio, the cost of selection, and the variability in criteria scores affect the economic utility of a selection device (Cascio, 1980; Schmidt et al., 1979). Cronbach and Gleser (1965) reached the same conclusions regarding the interaction of the above parameters. The equation below represents the Brogden, Cronbach and Gleser Utility model:

\[ U = N_s \cdot r_{xy} \cdot SD_y \cdot Z_x - N_s \cdot C/P \]

Where \( N_s \) = Number of selectees

\( C \) = Cost of testing one applicant

\( P \) = Selection ratio
Zx = Average standard score on the test of those selected (in applicant group standard score units)

rxy = Test validity

SDy = The standard deviation of job performance in dollar terms among randomly selected employees

The only assumption required by this formula, a linear relationship between test score and job performance, is widely supported in the research literature (Schmidt and Hunter, 1980; Schmidt, Hunter, Pearlman & Shane, 1979). As the formula indicates, even selection tests with low validities may have significant utility when the standard deviation of the criterion is large (Cascio, 1982). Unlike the Taylor and Russell model, utility is represented in dollars and may be compared across situations.

However, because of the difficulty and expense of obtaining criteria outcomes expressed in dollars through cost-accounting methods, this model has not been widely applied (Cascio, 1983; Schmidt & Hunter, 1980). In an attempt to overcome this problem, Schmidt, Hunter, McKenzie, and Muldrow (1979) developed a rational method for estimating the SD of job performance. Supervisors are asked to make estimates of the annual dollar value to the organization of a poor employee (performing at 15th
percentile), an average employee (performing at the 50th percentile), and a superior employee (performing at the 85th percentile). The difference between the estimates of the poor and average workers and the difference between the estimates of the average and superior workers are calculated. The average of these two differences is then calculated and designated as one standard deviation (SD) of the criterion in dollar terms. The reasoning on which this procedure is based is as follows: if job performance in dollar terms has a normal distribution, the difference between the values of the products and services produced by an average worker (50th percentile) and those produced by a poor worker (15th percentile) or superior worker (85th percentile) is equal to one SD of the criteria.

The rational method for estimating the SD was first utilized by Schmidt et al. (1979). The results indicated that hundreds of millions of dollars in increased productivity could be realized by increasing the validity of selection decisions in hiring government computer programmers. Weekley, Frank, O'Connor, and Peters (1983) indicated that only recently have researchers begun to examine the utility of various personnel selection procedures in dollar terms. Most applications of the model have indicated similar results, that is, even small increases in validity may result in large monetary savings, especially for larger organizations (Schmidt et
al., 1979; Schmidt, Mach, & Hunter, 1983; Schmidt & Hunter, 1983; Weekley, Frank, O'Connor, and Peters, 1983). In addition, Bobko, Karren, and Parkington (1983) reported evidence in support of the ability of supervisors and managers to estimate the standard deviation of the criterion in dollar terms. With the availability of the rational method for estimating the standard deviation of the criterion, Cascio (1982) suggested that "Utility analysis should become a central concept in personnel selection research" (p. 225). It seems equally advisable for organizations who are concerned with the costs and benefits associated with selection procedures to perform utility analyses.
CHAPTER II

THE PRESENT STUDY

The present study represents an attempt to develop, validate and estimate the utility of a WAB to be used as part of the screening process for selecting truck drivers in a midwestern division of an international delivery company (IDC). The IDC operates under a management by objective (MBO) program. As part of this program, the delivery personnel are given daily time-objectives. The time objectives cover every aspect of the worker's day from the time he/she starts work to the time he/she finishes. The time analyses were developed in two forms: one for driving activities and another for non-driving activities. The non-driving time estimates were developed by calculating the average amount of time needed by a worker to perform every activity that is required to complete the assigned tasks. For example, before delivering any products a worker must pre-trip or check his/her delivery vehicle. Time analyses were computed which estimated the amount of time required to perform this procedure. The procedure was broken down into small time elements such as time required to walk necessary distances, execute all turns, stoops, and observations. Every non-driving activity required of the worker for every delivery route is calculated in this manner. Extraneous variables which may alter these expected times (e.g., customers not having cash or checks ready for
payment, or nonoperational elevators) are subtracted daily from each drivers actual working time for performance evaluations.

The driving aspects of the job have been subject to similar time analyses. Based on the routes a driver will be covering on a particular day, a specific amount of on-road time is allotted as that driver's daily on-road time-objective. The IDC has conducted time analyses on every route it covers, resulting in time objectives which are statistically reliable measures. The time analyses are computed by deriving the average amount of time it takes for a specific type of IDC truck, while traveling at a specified speed, to safely stop and to then resume a specific speed. By knowing the number of traffic regulators, such as stop signs and traffic lights, and the type of speed zone in which these regulators are located, an estimate of the average time needed to drive a particular route can be computed. Time variations due to factors beyond the driver's control are subsequently subtracted from the driver's time-objectives for evaluative purposes (e.g., time lost in a traffic jam would be subtracted from the driver's on-road time).

A total driving and non-driving time objective which takes into account every aspect of the worker's day is assigned daily to each driver. Performance evaluations are made by comparing the worker's adjusted working time with the estimated or standard time derived from the time analyses. In
1984 the IDC agreed to the development of a weighted application blank in an attempt to enhance the selection of satisfactory delivery drivers as defined by the above criterion.

Based on the literature, it was concluded that a biographical inventory developed by the EA would be an appropriate method for selecting delivery drivers who would perform satisfactorily on the job. This decision was based primarily on research which indicated that biographical data has had a higher percentage of validity coefficients equal to or above the .50 level than have other predictors (Asher & Sciarrino, 1974). The subsequent decision to use an empirical rather than a rational approach for scoring biographical data was based on the three points noted earlier: (a) there is evidence that suggests that the EA may yield higher derivation and cross validities than the RA (Mitchell & Klimoski, 1982; Moomaw, et al., 1984), (b) the EA takes less time, and (c) it is less expensive. In addition, and more importantly, the working arrangement with IDC required that only data already on file, that is, information on the delivery driver applications, could be utilized in the project. Thus, the use of the RA was impractical. In order to obtain a group of questions with high internal consistency, many items would have to be developed and administered to the delivery personnel. This process would violate the agreement to use only archival
data.

The decision was that the empirical approach of England (1971) would be used for scoring the biographical data in the present study. That decision was based on current literature which states that in addition to being the most frequently cited of all empirical methods, the England approach has consistently yielded predictors with high validity coefficients (Pace & Schoenfeldt, 1977; Weiss, 1976; Mitchell & Klimoski, 1982).

Hypotheses

Based on the high validity estimates obtained using biodata in past research it was hypothesized that (a) significant derivation and cross-validities would be obtained, and (b) significant practical benefits in terms of correct placement of workers in the high and low criterion groups would result.
CHAPTER III
METHODOLOGY

The England Method

The England (1971) method was utilized in the present study to develop a weighted application blank for the IDC. The following section describes the England method.

First, the criterion used to differentiate desirable and undesirable employees must be identified. The most relevant, reliable and accessible measure should be used as the criterion (England, 1971). The importance of selecting a reliable criterion cannot be overstated. Muchinsky (1983) and Owens (1983) stressed that the entire study is dependent on the reliability of the criterion--that is, the less reliable the criterion, the lower the potential for identifying valid predictors.

After the criterion is selected, the next step is to divide the subjects (employees) into high and low criterion groups based on their performance. The high criterion group consists of employees that perform well on the criterion measure, and the low criterion group consists of employees that perform poorly on the criterion measure. Both groups are subdivided into derivation and hold-out groups. The high and low derivation groups are used to identify and weight biographical questions which differentiate between desirable and undesirable employees, i.e., to develop the
WAB. The high and low hold-out groups are subsequently used to determine whether the derived WAB is valid for a group other than the one used in its development (Owens, 1983). In order to avoid problems with exaggerated sampling error, England (1971) stated that no study using the EA should be undertaken with fewer than 75 people in both the original high and low criterion groups. In addition, England suggested that the derivation groups consist of approximately twice as many people as the hold-out groups.

The next step in the process is the selection of the application blank items to be analyzed. Owens, Glennon, and Albright (1962) demonstrated that the development of the items is critical in determining the reliability and in turn the validity of the predictor. Four general rules for the development of biodata questions were offered:

1. Brevity is desirable; there seems to be a negative relationship between length and validity of the question.
2. Whenever possible, numbers should be used to graduate and to define options or alternatives.
3. Either all response options or alternatives should be covered or an "escape" option should be provided.
4. Items, particularly item stems, should carry a neutral or a pleasant connotation for the respondent. (p. 330)
In addition to developing questions, there are several sources of items that have proven valid in past research (e.g., Owens, 1983; England, 1971). It is suggested that since many items are discarded because they do not differentiate between the criterion groups, as many items as possible, within reason, should be used in the original validation process (Owens, 1983). In addition to selecting the questions, response categories for each item must also be provided.

After the questions and response categories are in questionnaire format, the questionnaire is administered to the subjects. The high and low derivation group's responses are compared to determine if desirable employees respond differently than undesirable employees. The weighting technique used by England (1971) is called the vertical percent method and involves three steps. First, for each response category, the percentage of the low group indicating a particular response is subtracted from the percentage of the high group indicating the same response. Second, using a table derived by E. K. Strong (reprinted in England, 1971; and Guion, 1965), the resulting percentages are converted into net weights. Third, using the Strong tables, the net weights are converted to assigned weights with smaller, positive values. The third step is optional and is done simply to avoid
calculating errors. Finally, any question that does not differentiate between the high and low derivation groups is discarded (i.e., any item in which all responses are weighted the same).

Once the items have been weighted and the invalid questions discarded, the WAB is cross-validated on the high and low hold-out groups. This procedure is done by summing the weights of each response designated by the hold-out groups on the WAB. If the WAB is valid, and a positive relationship exists, the total scores for the high hold-out group will be greater than those of the low hold-out group. The greater the overlap in the distributions of the scores the less valid is the WAB.

A closely related procedure is that of setting a cut off score for selection purposes. The cut off score indicates the point below which the employer will not hire. The optimal cut off score is one which allows the maximum number of people in the high hold-out group to be hired, while rejecting the maximum number of people in the low hold-out group. This is accomplished by identifying the total WAB score which represents the largest index of differentiation between the high and low performing workers (England, 1971). When the cut off score is set the EA approach is complete.

Data Collection

Criterion measures and biographical data were
collected from May 1984 through July 1984 in Denver, Colorado, at the mid-western division of the IDC.

**Criterion Measure**

Job proficiency as measured by the percentage of on-job time over or under the weekly time-objective was the criterion. A test-retest reliability estimate using the Pearson product-moment correlation was computed with a sample of 187 delivery drivers. Each driver's daily time-objectives were averaged for a one week period and then correlated with a second weekly average after a one week interval. The reliability coefficient was .64. This reliability estimate was then stepped up using the Spearman Brown formula (Anastasi, 1982). The resulting reliability estimate was .958.

**Validation Samples**

The total sample consisted of 187 delivery drivers from the midwestern division of the IDC. All subjects (a) had been working for IDC at least one year but no longer than five years and (b) had biographical and performance data on file.

Criterion data were collected for each of the 187 drivers for a six month period beginning with August 1983 and ending with January 1984. Each driver was rank ordered according to performance. Approximately the top third of the drivers (n = 62) were identified as performing well (high group). The mean criterion score
was .956 and ranged from .881 to .998. The standard deviation was .025. Approximately the bottom third of the drivers (n = 56) were identified as performing poorly (low group). The criterion scores for the low group ranged from 1.038 to 1.207 with the mean being 1.084. The standard deviation was .038. The remaining drivers (n = 69) were identified as average (ranging from 1.002 to 1.014) and were not included in further analyses.

The cross-validation sample (n = 44) consisted of a random sample of twenty-two drivers from the high group and twenty-two from the low group. The derivation sample (n = 74) consisted of a random sample of 40 drivers from the high group and 34 drivers from the low group.

**Development of Biographical Items.**

Forty-six biographical items representing continuous or nominal variables in a multiple-choice format were developed from the existing application blank of the IDC. The suggestions of Owens et al. (1962) and Asher (1972) were followed in writing the multiple-choice items. The development of the biographical questions was undertaken primarily for coding purposes—That is, the subjects actually responded to the questions on the application blank, not to the biographical questions. In many cases the
questions on the application were open ended and many variations in responses were possible. As described later in this paper, biographical questions that resulted in ambiguous responses were dropped from the study. The biographical questions and corresponding options developed from the application blank are shown in Appendix A.

**Item Weights**

In order to assign weights to the biodata items, it was first necessary to determine the extent to which the high and low derivation groups differed in their responses to the biodata questions. During the data collection, it was discovered that two different versions of the IDC application blank had been used to record applicant data. As a result questions 7, 16, 17, 43, 44, and 46 were dropped from further analyses because all applicants could not respond in the same manner. In addition, data provided for questions 10, 11, 12, 18, 22, 38, 39, and 41 were ambiguous or grossly incomplete. It should be noted that the incomplete or ambiguous information was not a result of the biographical questions, but a result of applicants failing to properly complete the original application blank.

For the remaining twenty-four questions, the percentage of the low derivation group who chose a
response was subtracted from the percentage of the high derivation group who chose the same response. Strong's tables originally reported in Stead & Shartle (1940) and reprinted in England (1971) and Guion (1965) were then used to convert the percentage differences into net weights. Thus, each response category was assigned a net weight. A third and optional step in the weighting process, converting the net weights to positive assigned weights with small values (England, 1971), was omitted in order to retain the more sensitive net weights (Mitchell & Klimoski, 1982). As a result of the weighting process, questions 8, 13, 15, 25, 29, 31, 34, 36, and 45 were identified as not differentiating between the two derivation groups (i.e., all of the response options for the question were assigned the same weight) and were dropped from the study. In most cases these results were expected. For example, every one indicated being a U.S. citizen and being able to operate a clutch. Thus, the questions were of no value. However, the analyses of some of the other questions point to other explanations for the lack of differences between the two groups, such as the screening of applicants based on certain items. For example, in question 36, no one in the sample indicated having been involved in an accident as an operator of a commercial vehicle. Perhaps IDC had already screened out any
applicant who indicated having such an accident. Implications of such screening procedures are discussed later in this paper as they relate to the present study.

In addition to the aforementioned questions, question numbers 24 and 37 were also discarded because the information they provided was redundant with the information provided by questions 23 and 35, respectively. Thus, a total of 21 questions were identified as differentiating between the high and low derivation samples (see Appendix B).

Obtaining a WAB Score

After the derivation of the item weights, a WAB score was computed for each derivation subject by summing the net weights of each subject's responses to the twenty-one questions. WAB scores were also computed for the high and low hold-out groups in the same manner.

The WAB scores for the high-derivation sample ranged from -19 to 41 with a mean of 16.8 and a standard deviation of 14.58. The WAB scores for the low-derivation sample ranged from -35 to 16 with a mean of -5.68 and a standard deviation of 15.30. The WAB scores for the high hold-out group ranged from -22 to 28 with a mean of 5.1 and a standard deviation of 11.1. The WAB scores for the low hold-out group ranged from -37 to 27 with a mean of 4.27 and a standard deviation of 14.92.
Utility Analysis Measures

The equation used for the Utility analysis is as follows:

\[ U = N_s \cdot r_{xy} \cdot SD_y \cdot Z_x - N_s \cdot \frac{C}{P} \]

Where \( N_s \) = Number of selectees

\( C \) = cost of testing one applicant

\( P \) = Selection ratio

\( Z_x \) = Average standard score on the test of those selected (in applicant group standard score units)

\( r_{xy} \) = Test validity

\( SD_y \) = The standard deviation of job performance in dollar terms among randomly selected employees

Estimates of the standard deviation of the criterion in dollar terms were provided by three experienced delivery driver managers. Each manager was given a questionnaire with detailed instructions (see Appendix D) for estimating the standard deviation. The questionnaire was a modified version of the original questionnaire and instructions provided by Schmidt et al. (1979). All three questionnaires were returned in usable form. The mean difference between the poor worker and the average worker was $3,166.66. The mean difference between the average worker and the superior worker was $1,933.34. The average of these two
estimates was $2,550.00 and was designated as the standard deviation of the criterion in dollar terms.

Additional parameters of the equation were estimated by the personnel manager as follows: Cost of testing one applicant = $.35, Selection Ratio for 1984 = .10.
CHAPTER IV
RESULTS

Significance Tests

The derivation validity estimate was calculated by correlating the total WAB scores of the derivation sample with their corresponding criterion scores. The resulting correlation was -.566 and was statistically significant (p < .001).

The cross-validity coefficient was calculated by correlating the total WAB scores of the hold-out sample with their corresponding criterion scores. The cross-validity coefficient was -.127 which was not statistically significant (p > .05).

The amount of shrinkage, that is, the loss in observed validity from derivation to cross-validation, was assessed using a test of the difference between two independent samples (Hayes, 1973). The shrinkage between the two validity coefficients was significant (Z = 2.65; p < .05).

Utility Analysis

Given an estimated WAB validity of "0," a .10 selection ratio, a $.35 testing cost per applicant and a hiring rate of 40 delivery drivers per year, the utility of the WAB procedure over a one year period was estimated to be $-140.00.
Discussion

Hypothesis one, that significant derivation and cross validities would be obtained, was partially supported by the aforementioned results, that is, the derivation validity was statistically significant ($r = -0.56$, $p < 0.001$). More importantly however, the cross-validity was not significant ($r = -0.12$, $p > 0.05$). Thus, the second hypothesis, that significant practical benefits in terms of correct placement of workers in the high or low criterion groups would be obtained, was not supported.

Since a significant amount of shrinkage occurred during cross-validation, the weights assigned to the derivation sample were not applicable to the cross-validation group. Questions which differentiated between the high and low derivation sample did not differentiate as well between the high and low cross-validation sample. To examine more closely the differences in the manner in which the derivation and cross-derivation sample answered questions, the percentages of people in the high and low derivation group choosing a response were compared with the percentage of people in the high and low cross-validation group choosing the same response (see Appendix C). The two groups answered most of the questions in a vastly different manner. For example, on question number five 82.4% of the high derivation group chose option number one while only 42.1% of the low derivation group chose option number one. Conversely, only 45.5% of the high
cross-validation group chose option number one, while a higher
65.0% of the low cross-validation group chose option number
one. The manner of answering question number five was
reversed between the derivation and cross-validation groups.
This type of reversal which occurred on several questions
helps to explain why such a high degree of shrinkage occurred
during cross-validation.

Several methodological explanations for the large amount
of shrinkage also exist. First, about 70% of IDC's delivery
drivers are performing their jobs satisfactorily. This figure
was derived by examining the criterion scores of 187 drivers
in the present study. Almost three fourths of the drivers had
a criterion score equal to or better than IDC's minimal
standards. This situation causes a high degree of restriction
of range in the criterion which may result in large amounts of
shrinkage (Mitchell and Klimoski, 1982). Similarly, it is
highly probable that IDC used the original applications to
screen the applicants. As with any concurrent study, the
screening of applicants may cause the sample population to be
homogeneous and cause a large amount of restriction of range
in the predictor (Arvy, 1979). Restriction of range in either
the criterion or predictor would deflate the observed validity
estimates (Mumford et al., 1984).

A related issue is that of training. The biographical
questions may not have been important with respect to the
prediction of differential performance once adequate training
was acquired (Mumford et al., 1984). This is a highly probable assumption since the IDC has an extensive training school which all delivery drivers are required to complete prior to assuming regular duties. This would also contribute to restriction of range in the criterion and predictor since those that do not perform satisfactorily in training are not retained by the company.

Another area of concern is the sample size used in the present study (N = 118). The sample size was slightly less than the minimal suggested sample size (N = 150) when using the England procedure (England, 1971). As Mitchell and Klimoski (1982) noted, WAB studies based on minimal sample sizes are especially prone to large amounts of shrinkage during cross-validation due to exaggerated sampling error (England, 1971).

Finally, the working arrangement with IDC required the study to be archival in nature--that is, only data that existed on the original application blanks were available for the study. The biodata were obtained by examining the completed applications of current employees. This procedure precluded the use of biographical items from existing item pools (e.g., Glennon et al., 1966) which have demonstrated validity in past studies. Also precluded was the development of original biodata questions to tap specific life history areas of delivery drivers (England, 1971). The application blank used by IDC was not constructed for conducting
biographical research. As such, it did not lend itself to obtaining the most appropriate biodata. Schmidt et al. (1979) noted that indices of past activity are likely to yield significant correlations with criteria measures only when the effectiveness of the past activity is examined. For example, in the present study, prior driving experience was examined, but no indication of the quality of the driving experience was available from the archival data. In addition, only 46 biodata questions could be answered from the application. Mitchell and Klimoski (1982) suggested that if the original question pool is small and/or poorly chosen large amounts of shrinkage may result during cross-validation. The results of the present study tend to support this statement.

Implications
From a legal and practical standpoint, the WAB developed in the present study should not be implemented. If challenged, the WAB would not be looked upon by the courts as a valid selection device because the cross-validation estimate is not significant. In addition, the utility of the WAB procedure is estimated to be negative because of its lack of validity and the costs associated with its implementation. Thus, the WAB has no legal or practical advantages associated with its implementation.

However, many results from the study are of practical value. As mentioned previously, the criterion, a ratio of planned work time to actual work time, was found to be
reliable. As reported, the test-retest reliability estimate for the criterion was .95. This should be taken as welcome news by IDC since the measure is utilized in many management functions.

The present study also revealed that, on the aggregate level, the questions used in the application apparently are not significantly related to job performance as measured by the ratio of planned work time to actual work time. This relationship creates a concern from a legal standpoint, that is, under certain conditions (i.e., when hiring decisions are based on application information, an EEO suit is filed by a rejected applicant in a protected class, and adverse impact exists) the IDC may not prevail in a legal defense of its hiring practices. It is recommended that IDC review the hiring process as it relates to the use of the material in the application. If use is made of the application information for hiring purposes, an attempt to validate the use of the material is recommended.

The results of the present study are also of value from a theoretical viewpoint. First, the results identify many problems associated with using archival data as the primary source for the development of a WAB. While nothing is inherently wrong with archival data, its use in WAB development seems to be limited if the original documents were not designed with the purpose of collecting biographical data.

Schwab and Oliver (1974) have commented that the
usefulness of the WAB may be overstated in the literature. Many biographical studies may not have used cross-validation during earlier years. The results of this study indicate that to the extent that capitalization on chance occurs, a proportionate decrease in cross-validity will follow. This study reiterates and mandates the utilization of cross-validation. Contrary to Schwab and Oliver's statement, and as indicated in this paper, there is now an extensive literature base which indicates that many biographical studies employing cross-validation yield high and significant validity estimates.

Future WAB studies should be conducted with the awareness of the problems outlined above associated with restriction of range, small or minimal sample sizes, and poorly chosen and/or small item pools. In addition, it is recommended that items which tap specific life history areas be developed (England, 1971).
1. In how many different locations have you resided during the last three years?
   1. one
   2. two
   3. three

2. Have you resided in a state other than Colorado within the last three years?
   1. yes
   2. no

3. Would you accept night work?
   1. yes
   2. no

4. Are you now employed?
   1. yes
   2. no

5. How much advanced notice time (in days) would you require before starting work with IDC?
   1. zero, can start immediately
   2. one to seven days
   3. eight days or more

6. Have you ever applied at IDC before?
   1. yes
   2. no

7. Why are you applying for a job at IDC?
   1. high wages & benefits
   2. job security
   3. other

8. Are you a U.S. citizen?
   1. yes
   2. no

9. How many years of formal education have you completed?
   1. 12 or under
   2. 13 or 14
   3. 15 or more

10. How many years of college have you completed?
    1. zero
    2. one to two
    3. three to four
    4. five or more
11. How many years of technical or business school have you completed?
   1. zero
   2. one to two
   3. three to four
   4. five or more

12. If you attended any school beyond high school, what was your major area of study?
   1. business
   2. liberal studies
   3. education
   4. other

13. Did you graduate from high school?
   1. yes
   2. no

14. Did you graduate from college?
   1. yes
   2. no

15. Did you graduate from a technical or business school?
   1. yes
   2. no

16. How many jobs (other than military) have you held prior to filling out this application?
   1. zero
   2. one to three
   3. four to six
   4. six or more.

17. What was your hourly pay for your most recent job?
   1. 3.50 or under
   2. 3.51 to 5.50
   3. 5.51 or over

18. What was your reason for leaving your most recent previous job?
   1. does not apply
   2. moved
   3. layed off
   4. other

19. Do you wish us not to contact any of your previous employers?
   1. Yes
   2. No

20. How many times have you been discharged from previous employment? (other than military)
   1. never
   2. once or more
21. How many weeks have you collected Unemployment compensation during the last five years?
   1. never, does not apply
   2. one to 10 weeks
   3. eleven to twenty weeks
   4. twenty-one weeks or more

22. In what branch of the military have you served?
   1. does not apply
   2. Army
   3. Mavy
   4. Air Force
   5. Marines
   6. other

23. How many years have you been driving a vehicle?
   1. none
   2. 1 to 3
   3. 4 to 6
   4. 7 or more

24. How many years have you been driving commercially?
   1. none
   2. 1 to 3
   4. 4 or more

25. Can you drive a clutch operated vehicle?
   1. yes
   2. no

26. In how many different states have you held a driving license?
   1. zero or one
   2. two or more

27. How many restrictions do you have on your drivers license?
   1. zero
   2. one or more

28. Has any license you have ever held been suspended?
   1. yes
   2. no

29. Has any license you have ever held been revoked?
   1. yes
   2. no

30. How many years of driving experience do you have with tractor-trailers?
   1. none
   2. 1 or more
31. How many years of driving experience do you have with buses?
   1. none
   2. 1 or more

32. How many years of driving experience do you have with vans?
   1. none
   2. 1 or more

33. How many years of driving experience do you have with 2 1/2 ton trucks?
   1. none
   2. 1 or more

34. How many safe driving awards have you been awarded?
   1. zero
   2. one or more

35. How many accidents have you been involved in, regardless of severity?
   1. zero
   2. one or more

36. How many accidents have you been involved in as an operator of a commercial vehicle?
   1. zero
   2. one or more

37. How many accidents have you been involved in as an operator of a private vehicle?
   1. zero
   2. one or more

38. How many accidents have you had in the past year?
   1. zero
   2. one or more

39. How many accidents have you had in the past two years?
   a. zero
   b. one or more

40. How many traffic violations, other than parking, have you been convicted of?
   1. zero
   2. one or two
   3. three or four
   4. four or more

42. What types of traffic violations, other than parking, have you ever been convicted of?
   1. does not apply
   2. speeding
   3. other
43. Do you wear glasses?
   1. yes
   2. no

44. Do you wear contact lenses?
   1. yes
   2. no

45. Are you color blind?
   1. yes
   2. no

46. If you wear glasses or contact lenses is your vision in each eye at least 20/40 with prescription?
   1. yes
   2. no
   3. does not apply
APPENDIX B

BIOGRAPHICAL ITEMS & RESPONSES WHICH DIFFERENTIATED BETWEEN DELIVERY PERSONNEL IN THE DERIVATION SAMPLE

1. In how many different locations have you resided during the last three years?
   1. one
   2. two
   3. three

2. Have you resided in a state other than Colorado within the last three years?
   1. yes
   2. no

3. Would you accept night work?
   1. yes
   2. no

4. Are you now employed?
   1. yes
   2. no

5. How much advanced notice time (in days) would you require before starting work with IDC?
   1. zero, can start immediately
   2. one to seven days
   3. eight days or more

6. Have you ever applied at IDC before?
   1. yes
   2. no

9. How many years of formal education have you completed?
   1. 12 or under
   2. 13 or 14
   3. 15 or more

14. Did you graduate from college?
   1. yes
   2. no

19. Do you wish us not to contact any of your previous employers?
   1. Yes
   2. No
20. How many times have you been discharged from previous employment? (other than military)
   1. never
   2. once or more

21. How many weeks have you collected Unemployment compensation during the last five years?
   1. never, does not apply
   2. one to 10 weeks
   3. eleven to twenty weeks
   4. twenty-one weeks or more

23. How many years have you been driving a vehicle?
   1. none
   2. 1 to 3
   3. 4 to 6
   4. 7 or more

26. In how many different states have you held a driving license?
   1. zero or one
   2. two or more

27. How many restrictions do you have on your driver's license?
   1. zero
   2. one or more

28. Has any license you have ever held been suspended?
   1. yes
   2. no

30. How many years of driving experience do you have with tractor-trailers?
   1. none
   2. 1 or more

32. How many years of driving experience do you have with vans?
   1. none
   2. 1 or more

33. How many years of driving experience do you have with 2 1/2 ton trucks?
   1. none
   2. 1 or more
35. How many accidents have you been involved in, regardless of severity?
   1. zero
   2. one or more

40. How many traffic violations, other than parking, have you been convicted of?
   1. zero
   2. one or two
   3. three or four
   4. four or more

42. What types of traffic violations, other than parking, have you ever been convicted of?
   1. does not apply
   2. speeding
   3. other
### Derivation and Cross-validation Group Percentages

#### Derivation Group

<table>
<thead>
<tr>
<th>Biodata Questions</th>
<th>HG</th>
<th>LG</th>
<th>DIFF</th>
<th>WT</th>
<th>HG</th>
<th>LG</th>
<th>DIFF</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38.2%</td>
<td>52.5%</td>
<td>-14.3%</td>
<td>-3</td>
</tr>
<tr>
<td>option 1</td>
<td>38.2%</td>
<td>30.0</td>
<td>8.2</td>
<td>2</td>
<td>40.9%</td>
<td>68.2%</td>
<td>-34.7%</td>
<td>-6</td>
</tr>
<tr>
<td>option 2</td>
<td>23.5</td>
<td>17.5</td>
<td>6.0</td>
<td>1</td>
<td>18.2</td>
<td>13.6</td>
<td>4.6</td>
<td>1</td>
</tr>
<tr>
<td>option 3</td>
<td>26.5</td>
<td>43.6</td>
<td>-17.1</td>
<td>-4</td>
<td>36.4</td>
<td>22.7</td>
<td>13.7</td>
<td>3</td>
</tr>
<tr>
<td>option 2</td>
<td>73.5</td>
<td>56.4</td>
<td>10.2</td>
<td>4</td>
<td>63.6</td>
<td>77.3</td>
<td>-13.7</td>
<td>-3</td>
</tr>
<tr>
<td>#2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>84.8</td>
<td>95.0</td>
<td>10.2</td>
<td>-3</td>
</tr>
<tr>
<td>option 1</td>
<td>15.2</td>
<td>5.0</td>
<td>-10.2</td>
<td>3</td>
<td>13.6</td>
<td>14.3</td>
<td>-0.7</td>
<td>0</td>
</tr>
<tr>
<td>option 2</td>
<td>73.5</td>
<td>56.4</td>
<td>11.0</td>
<td>2</td>
<td>63.6</td>
<td>86.4</td>
<td>-22.8</td>
<td>-5</td>
</tr>
<tr>
<td>option 2</td>
<td>26.5</td>
<td>37.5</td>
<td>-11.0</td>
<td>-2</td>
<td>36.4</td>
<td>13.6</td>
<td>22.8</td>
<td>5</td>
</tr>
<tr>
<td>#3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73.5</td>
<td>56.4</td>
<td>11.0</td>
<td>2</td>
</tr>
<tr>
<td>option 1</td>
<td>29.5</td>
<td>17.5</td>
<td>12.0</td>
<td>3</td>
<td>21.8</td>
<td>22.6</td>
<td>-0.8</td>
<td>0</td>
</tr>
<tr>
<td>option 2</td>
<td>82.4</td>
<td>42.1</td>
<td>40.3</td>
<td>10</td>
<td>45.5</td>
<td>65.0</td>
<td>-19.5</td>
<td>-4</td>
</tr>
<tr>
<td>option 2</td>
<td>17.6</td>
<td>57.9</td>
<td>-40.3</td>
<td>-10</td>
<td>45.4</td>
<td>35.0</td>
<td>10.4</td>
<td>4</td>
</tr>
<tr>
<td>#4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39.4</td>
<td>27.5</td>
<td>11.9</td>
<td>3</td>
</tr>
<tr>
<td>option 1</td>
<td>60.6</td>
<td>72.5</td>
<td>-11.9</td>
<td>-3</td>
<td>72.7</td>
<td>68.2</td>
<td>4.5</td>
<td>1</td>
</tr>
<tr>
<td>option 2</td>
<td>32.4</td>
<td>55.0</td>
<td>-22.6</td>
<td>-5</td>
<td>40.9</td>
<td>18.2</td>
<td>22.7</td>
<td>5</td>
</tr>
<tr>
<td>option 2</td>
<td>38.2</td>
<td>27.5</td>
<td>10.5</td>
<td>2</td>
<td>27.3</td>
<td>59.1</td>
<td>-31.8</td>
<td>-7</td>
</tr>
<tr>
<td>option 3</td>
<td>29.5</td>
<td>17.5</td>
<td>12.0</td>
<td>3</td>
<td>21.8</td>
<td>22.6</td>
<td>-0.8</td>
<td>0</td>
</tr>
<tr>
<td>#9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.1</td>
<td>5.3</td>
<td>3.8</td>
<td>2</td>
</tr>
<tr>
<td>option 1</td>
<td>9.1</td>
<td>5.3</td>
<td>3.8</td>
<td>2</td>
<td>9.1</td>
<td>4.6</td>
<td>4.5</td>
<td>1</td>
</tr>
<tr>
<td>option 2</td>
<td>90.9</td>
<td>94.7</td>
<td>-3.8</td>
<td>-2</td>
<td>86.4</td>
<td>90.9</td>
<td>-4.5</td>
<td>-1</td>
</tr>
<tr>
<td>#14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81.8</td>
<td>97.5</td>
<td>-15.7</td>
<td>-5</td>
</tr>
<tr>
<td>option 1</td>
<td>81.8</td>
<td>97.5</td>
<td>-15.7</td>
<td>-5</td>
<td>90.9</td>
<td>90.9</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>option 2</td>
<td>18.2</td>
<td>2.5</td>
<td>15.7</td>
<td>5</td>
<td>9.1</td>
<td>9.1</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>#20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>78.8</td>
<td>65.0</td>
<td>13.8</td>
<td>3</td>
</tr>
<tr>
<td>option 1</td>
<td>21.2</td>
<td>35.0</td>
<td>-13.8</td>
<td>-3</td>
<td>9.0</td>
<td>22.7</td>
<td>-13.7</td>
<td>-3</td>
</tr>
<tr>
<td>option 2</td>
<td>36.7</td>
<td>16.2</td>
<td>20.5</td>
<td>5</td>
<td>14.3</td>
<td>9.5</td>
<td>4.8</td>
<td>1</td>
</tr>
<tr>
<td>option 2</td>
<td>13.3</td>
<td>32.4</td>
<td>-19.1</td>
<td>-4</td>
<td>28.6</td>
<td>28.6</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>option 3</td>
<td>26.7</td>
<td>29.7</td>
<td>-3.0</td>
<td>-1</td>
<td>28.6</td>
<td>19.0</td>
<td>9.6</td>
<td>2</td>
</tr>
<tr>
<td>option 4</td>
<td>23.4</td>
<td>21.6</td>
<td>1.8</td>
<td>0</td>
<td>28.6</td>
<td>42.9</td>
<td>-14.3</td>
<td>-3</td>
</tr>
<tr>
<td>#26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77.4</td>
<td>60.5</td>
<td>16.9</td>
<td>4</td>
</tr>
<tr>
<td>option 1</td>
<td>71.4</td>
<td>42.9</td>
<td>28.5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>option 2</td>
<td>28.6</td>
<td>57.1</td>
<td>-28.5</td>
<td>-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biodata Questions</td>
<td>Derivation Group</td>
<td>Cross-validation group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HG</td>
<td>LG</td>
<td>DIFF</td>
<td>WT</td>
<td>HG</td>
<td>LG</td>
<td>DIFF</td>
<td>WT</td>
</tr>
<tr>
<td>#27 option 1</td>
<td>77.4</td>
<td>89.5</td>
<td>-12.1</td>
<td>-3</td>
<td>81.8</td>
<td>81.0</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>option 2</td>
<td>22.6</td>
<td>10.5</td>
<td>12.1</td>
<td>3</td>
<td>13.6</td>
<td>19.0</td>
<td>-5.4</td>
<td>-1</td>
</tr>
<tr>
<td>#28 option 1</td>
<td>12.5</td>
<td>28.9</td>
<td>-16.4</td>
<td>-4</td>
<td>9.1</td>
<td>9.5</td>
<td>-0.4</td>
<td>0</td>
</tr>
<tr>
<td>option 2</td>
<td>87.5</td>
<td>71.1</td>
<td>16.4</td>
<td>4</td>
<td>90.9</td>
<td>90.5</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>#30 option 1</td>
<td>90.3</td>
<td>86.8</td>
<td>3.5</td>
<td>1</td>
<td>81.8</td>
<td>85.0</td>
<td>-3.2</td>
<td>1</td>
</tr>
<tr>
<td>option 2</td>
<td>9.7</td>
<td>13.2</td>
<td>-3.5</td>
<td>-1</td>
<td>13.6</td>
<td>15.0</td>
<td>-1.4</td>
<td>0</td>
</tr>
<tr>
<td>#32 option 1</td>
<td>84.8</td>
<td>92.1</td>
<td>-7.7</td>
<td>-2</td>
<td>77.3</td>
<td>85.0</td>
<td>-7.7</td>
<td>-2</td>
</tr>
<tr>
<td>option 2</td>
<td>15.5</td>
<td>7.9</td>
<td>7.6</td>
<td>2</td>
<td>22.8</td>
<td>15.0</td>
<td>7.8</td>
<td>2</td>
</tr>
<tr>
<td>#33 option 1</td>
<td>75.0</td>
<td>70.3</td>
<td>4.7</td>
<td>1</td>
<td>72.7</td>
<td>70.0</td>
<td>2.7</td>
<td>1</td>
</tr>
<tr>
<td>option 2</td>
<td>25.0</td>
<td>29.7</td>
<td>-4.7</td>
<td>-1</td>
<td>18.2</td>
<td>30.0</td>
<td>-11.8</td>
<td>-3</td>
</tr>
<tr>
<td>#35 option 1</td>
<td>43.8</td>
<td>34.2</td>
<td>9.6</td>
<td>2</td>
<td>50.0</td>
<td>28.6</td>
<td>21.4</td>
<td>5</td>
</tr>
<tr>
<td>option 2</td>
<td>34.4</td>
<td>42.1</td>
<td>-7.7</td>
<td>-2</td>
<td>27.3</td>
<td>23.8</td>
<td>3.5</td>
<td>1</td>
</tr>
<tr>
<td>option 3</td>
<td>21.9</td>
<td>23.6</td>
<td>-1.7</td>
<td>0</td>
<td>22.7</td>
<td>47.6</td>
<td>-24.9</td>
<td>-6</td>
</tr>
<tr>
<td>#40 option 1</td>
<td>18.8</td>
<td>16.2</td>
<td>2.6</td>
<td>1</td>
<td>22.7</td>
<td>14.0</td>
<td>8.7</td>
<td>2</td>
</tr>
<tr>
<td>option 2</td>
<td>34.4</td>
<td>32.4</td>
<td>2.0</td>
<td>0</td>
<td>31.8</td>
<td>19.0</td>
<td>12.8</td>
<td>3</td>
</tr>
<tr>
<td>option 3</td>
<td>21.9</td>
<td>18.9</td>
<td>3.0</td>
<td>1</td>
<td>27.3</td>
<td>23.8</td>
<td>3.5</td>
<td>1</td>
</tr>
<tr>
<td>option 4</td>
<td>25.0</td>
<td>32.4</td>
<td>-7.4</td>
<td>-2</td>
<td>18.1</td>
<td>42.9</td>
<td>-24.8</td>
<td>-6</td>
</tr>
<tr>
<td>#42 option 1</td>
<td>18.8</td>
<td>13.2</td>
<td>5.6</td>
<td>2</td>
<td>23.8</td>
<td>14.3</td>
<td>9.5</td>
<td>2</td>
</tr>
<tr>
<td>option 2</td>
<td>50.0</td>
<td>50.0</td>
<td>0.0</td>
<td>-</td>
<td>42.9</td>
<td>52.4</td>
<td>-9.5</td>
<td>-2</td>
</tr>
<tr>
<td>option 3</td>
<td>31.3</td>
<td>36.9</td>
<td>-5.6</td>
<td>-</td>
<td>33.3</td>
<td>33.2</td>
<td>0.1</td>
<td>0</td>
</tr>
</tbody>
</table>

Legend: HG = High Group  
LG = Low Group  
DIFF = Percent Difference between HG and LG  
APPENDIX D

UTILITY ESTIMATE INSTRUCTIONS

The dollar utility estimates we are asking you to make are critical in estimating the relative dollar value to IDC of different selection methods for delivery drivers. In answering these questions, you will have to make some very difficult judgments. We realize they are difficult and that they are judgments or estimates. You will have to ponder for some time before giving each estimate, and there is probably no way you can be absolutely certain your estimate is accurate when you do reach a decision. But keep in mind three things:

(1) The alternative to estimates of this kind is the application of cost accounting procedures to the evaluation of job performance. Such applications are usually prohibitively expensive. And in the end, they produce only imperfect estimates, like this estimation procedure.

(2) Your estimates will be averaged in with those of other supervisors of delivery drivers. Thus errors produced by too high and too low estimates will tend to be averaged out, providing more accurate final estimates.

(3) The decisions that must be made about selection methods do not require that all estimates be accurate down to the last dollar. Substantially accurate estimates will lead to the same decisions as perfectly accurate estimates.
Based on your experience with IDC delivery drivers, we would like for you to estimate the yearly value to IDC of the services produced by the average delivery driver. Consider the quality and quantity of output typical of the average delivery driver and the value on this output.

Based on my experience, I estimate the value to IDC of the average delivery driver at ( ) dollars per year.

We would now like for you to consider the "superior" delivery driver. Let us define a superior delivery driver as one who is at the 85th percentile. That is, his or her performance is better than that of 85% of his or her fellow delivery drivers, and only 15% turn in better performances. Consider the quality and quantity of the output typical of the superior delivery driver. Then estimate the value of these products and services.

Based on my experience, I estimate the value to IDC of a superior delivery driver to be ( ) dollars per year.

Finally, we would like you to consider the "low performing" delivery driver. Let us define a low performing deliver driver as one who is at the 15th percentile. That is, 85% of all delivery drivers turn in performances better than the low performing deliver driver, and only 15% turn in worse performances. Consider the quality and quantity of the output typical of the low performing delivery driver. Then estimate the value of these services.

Based on my experience, I estimate the value to IDC of the low performing delivery driver at ( ) dollars per year.
References


