PREDICTIVE VALIDITY OF THE RETAIL EMPLOYMENT INVENTORY
FOR THE SELECTION OF OVER-THE-ROAD TRUCK DRIVERS

THESIS

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Richard Edwin Nygren, Jr., B.A.
Denton, Texas
May, 1988
Nygren, Richard E. Predictive Validity of the Retail Employment Inventory for the Selection of Over-the-Road Truck Drivers. Master of Science (Industrial/Organizational Psychology), May, 1988, 47 pp., 7 tables, bibliography, 30 titles.

An independent pilot study suggested that the Retail Employment Inventory (REI) might be predictive of subjectively and objectively measured Over the Road (OTR) truck driver performance. The present validation study consisted of three parts.

First, an examination of the relationship between REI scores and 11 objective, performance criteria revealed weak and non-significant correlations. Second, a comparison of subjective ratings and REI scores failed to replicate the findings of the pilot study. And third, to confirm that the task components of the OTR job were correctly identified in the pilot study, a second job analysis was performed.

Possible reasons for the failure of the REI to predict OTR performance and directions for future research are discussed.
# TABLE OF CONTENTS

PREDICTIVE VALIDITY OF THE RETAIL EMPLOYMENT INVENTORY FOR THE SELECTION OF OVER THE ROAD TRUCK DRIVERS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Method</td>
<td>11</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td></td>
</tr>
<tr>
<td>Procedure</td>
<td></td>
</tr>
<tr>
<td>Criterion Measures</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>17</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Appendix</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>48</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Descriptive Statistics on All Variables.</td>
<td>18</td>
</tr>
<tr>
<td>2.</td>
<td>Correlation Coefficients Computed Between REI Scores and Subjective Performance Ratings.</td>
<td>20</td>
</tr>
<tr>
<td>3.</td>
<td>Correlation Coefficients Computed Between REI Scores and Objective Performance Ratings.</td>
<td>21</td>
</tr>
<tr>
<td>4.</td>
<td>Correlation Coefficients Computed Between Driver's Subjective Performance Ratings and Objective Performance Ratings.</td>
<td>22</td>
</tr>
<tr>
<td>5.</td>
<td>Intercorrelations Between Objective Variables</td>
<td>23</td>
</tr>
<tr>
<td>6.</td>
<td>Intercorrelations Between Subjective Rating Questions.</td>
<td>25</td>
</tr>
<tr>
<td>7.</td>
<td>Frequency Count of Job Related Characteristics Selected by SME's as Being Most Important to OTR Proficiency.</td>
<td>46</td>
</tr>
</tbody>
</table>
PREDICTIVE VALIDITY OF THE RETAIL EMPLOYMENT INVENTORY
FOR THE SELECTION OF OVER THE ROAD TRUCK DRIVERS

Since the beginning of intelligent and purposeful human activity, people have consistently sought out the best and most efficient way to get a job or task done. This goal was sometimes accomplished by selecting the best tool, by selecting the most efficient method, or by selecting the right person to do the job. This paper is concerned with the task of selecting the best people to perform the job of Over-the-Road (OTR) truck driving. Prior to discussing the design and findings concerning this endeavor, a review of the topic of personnel selection is in order.

The idea of using tests to select and classify people for jobs formally emerged through the work of Hugo Munsterberg. In his book *Psychology and Industrial Efficiency* (Munsterberg, 1913), he addressed the issues of selecting workers, designing work situations, and using psychology in sales (Muchinski, 1983). The work of Munsterberg, and that of other pioneering Industrial/Organizational (I/O) psychologists, marked the formal rapprochement of Psychology and Industry in the United States.

The outbreak of World War I coincided with the advent of the use of psychological tests to better understand peoples' interests and abilities, and then APA president Robert Yerkes suggested that psychologists' role in the war
effort might be a screening of recruits for mental deficiencies and assigning them to jobs within the Army (Muchinski, 1983).

This application of psychology emerged from the first World War as an acceptable and potentially valuable tool for industry. The large scale testing of both soldiers and industrial workers provided methodology, respectability, and stimulation to both academic and industrial thought about the potential utility of tests in the assessment of occupational aptitude (Ghiselli, 1973).

By 1941 and the outbreak of World War II, the profession of Industrial/Organizational Psychology and the science of personnel selection in particular had matured considerably from its infancy just prior to World War I. Throughout the war, there was a great increase in the use of employment tests in industry to increase productivity and reduce absenteeism (Pickard, 1945). American Industry had discovered that it could use the techniques of I/O Psychology developed for the military to better select, motivate, satisfy, and retain their employees. In fact, some have said that World War II was the springboard on which the techniques of Industrial Psychology were refined and catapulted into the limelight of American Industry (Muchinski, 1983).

After World War II, research in the area of personnel selection and testing continued. Much of this research
focused on the relationship between predictors and criteria, and an abundance of research has been undertaken in this area (Ghiselli & Haire, 1960; Guion & Cranny, 1982; Lawshe, 1985).

With regard to the selection of Over The Road (OTR) truck drivers, however, there appears to be a poverty of research. Some research, however, has been done in the area of selecting and classifying drivers for automobiles and light to medium trucks.

The earliest efforts were those of Robert N. McMurray and the White Motor Company in their "Continuing Control System of Truck Management." Between 1943 and 1950, McMurray developed a selection system called "Driver Selection and Training Forms" using a series of hurdles including an employment application for city drivers and one for long distance drivers, a telephone check form to be used when checking references, a patterned interview form to guide the interviewer, and supplementary but optional tests of traffic and driver knowledge including a road test in traffic. McMurray's system was heralded as "... a remarkable and generally sound program for selection and diagnosis in training..." (Wallace, 1953, p. 789); however, aside from imposing some degree of order on a previously unorganized selection scheme, McMurray's system falls well short of modern testing standards. Little information was provided by McMurray regarding the
validation of his tests and it was noted that there was no data on reliability or validity, no manual, and no normative description of the sample of 108 drivers.

Once order had been imposed on the selection of truck drivers in terms of interview and application forms, research efforts were then focused on identifying the attitudes which differentiated poor from good drivers. Between 1941 and 1958, Elmer B. Siebrecht developed the Siebrecht Attitude Scale which was purported to measure attitudes about safe driving practices. Even though Siebrecht’s scale has been of some interest to researchers, this reviewer could find no information on the scale’s reliability, validity, or normative data. In spite of Siebrecht’s attempts to identify unsafe driving attitudes, subsequent research employing his attitude scale casts serious doubt on its ability to differentiate accident from non-accident populations.

Forlano and Wrightstone (1956) attempted to determine the relationship between driver attitude and such factors as certain aspects of personality, driver knowledge, and intelligence. They examined scores on the Siebrecht Attitude Scale, the Bell Adjustment Inventory, the Standard Driver Knowledge Test, and the Henmon-Nelson Test of Mental Ability for a population of 1,352 secondary school students. They concluded that:
Based on the groups and tests utilized, the indications are that expressed driver attitude (e.g., toward driving violations, condition of drivers, courtesy, knowledge, and skills) is associated most with intelligence, then driver knowledge, and least of all with certain aspects of personality adjustments such as home, health, and social and emotional adjustment as measured by a self-descriptive inventory. (Forlano & Wrightstone, 1956, p. 186)

In a subsequent study of the reliability and the validity of the Siebrecht Attitude Scale (SAS), Crawford (1968) administered the SAS to two groups, one composed of individuals involved in two or more recent, at-fault accidents and the other composed of non-accident individuals. He concluded that:

... there was no significant difference between the two groups on any of the matching variables ... (and) ... that these findings cast serious doubt on the usefulness of this scale (SAS) in differentiating accidents from non-accident groups of drivers, or for making predictions about future driver behavior of employees or applicants for driver positions ... ." (Crawford, 1968, pp. 611-612)

A number of tests designed to select safe drivers were developed during the 1960's. However, most of these tests have been criticized for their lack of construct and
predictive validity. The Truck Driver Test, developed by the Public Personnel Association (Buros, 1965), is primarily an information test and its manual is devoid of actual validity evidence (Kerr, 1965). The McGuire Safe Driver Scale and Interview Guide (Buros, 1965) was designed to help select the safe driver; however, it has been criticized in that no validity coefficients are given and there is no validity evidence in pre-employment evaluation situations (Kerr, 1972). The Columbia Driver Judgment Test, developed by Fine, Malfetti, and Shoben (1965), was designed "... for use in driver selection and training, and as a criterion measure for driver behavior research ..."; but, the test has been criticized (Shuster, 1972) for not defining what good and bad drivers are and what driver behaviors it purports to predict (e.g., moving violations, accidents, etc.).

In an interesting study titled "Psychology of Drivers in Traffic Accidents," Preston and Harris (1965) compared 50 automobile drivers involved in at-fault accidents serious enough to require hospitalization to 50 drivers without accident histories. The subjects were compared on the basis of their driving experience and performance on the Rosenzweig Picture Frustration Study and Siebrecht Attitude Scale. The accident group differed from the safe group in their driving history but were indistinguishable on the written tests. The accident group described themselves as
much closer to expert than very poor on a driving performance continuum and "... in fixing the responsibility for the accidents and in estimating their driving competence at the time of the accidents, the accident subjects reports were at considerable variance with police reports" (Preston & Harris, 1965; p. 284).

Ghiselli (1973) determined that, with regard to motor vehicle operators (e.g., taxi cab drivers and bus drivers), tests of motor abilities (tapping, dotting, finger dexterity, and hand dexterity) and of personality traits (interests) seem to be best in predicting job proficiency, while tests of intellectual ability (intelligence and immediate memory) and perceptual accuracy (number comparison, name comparison, and perceptual speed) had little or no validity in the prediction of job proficiency.

This collection of research and minimally effective selection tests, then, would appear to stand alone in a field where there is a real need for improved selection strategies. With regard to the current OTR situation in the United States, large national (48 states) carriers are faced with: a) a shortage of qualified drivers, b) driver turnover rates averaging around 140 percent per year, and c) a public perception that killer trucks on our highway system are endangering the safety of the average motorist (Clearly, 1987). It is this setting, then in which freight carriers must make their selection decisions. The need for a test
which would help carriers identify and screen out applicants who are likely to engage in unsafe driving or rapid turnover is clearly evident.

The idea of using tests to help improve the selection of applicants is appealing to OTR operations; however, the abundance of questionably effective instruments described previously seems to have created a sense of uncertainty within carriers regarding the use of tests in their selection strategies (A.D. Thompson, personal communication, August, 1987). Smaller trucking companies are forced to continue to select drivers as they have in the past (application, reference checks, interview, etc.) while larger carriers, if there is a commitment to selection research, are put in the position of researching available selection instruments developed for other industries and attempting to validate them in the OTR setting (A.D. Thompson, personal communication, June, 1987).

The present research was conducted within a large, national 48 state carrier operating approximately 1,800 OTR trucks. A selection test called the Retail Employment Inventory (REI) was chosen by the human resources department of this large national carrier to be validated in a predictive, uncontaminated, criterion related study with OTR truck drivers.

The REI was selected by this carrier because the underlying constructs of this test, the measurement of
attitudes about responsible, dependable, and conscientious behavior, were thought to be particularly important to OTR proficiency. It was posited, then, that people who hold attitudes supporting responsible, dependable, and conscientious behavior would in fact perform their OTR job in a dependable, responsible, and conscientious manner. It was also thought that people who endorse these values might be less likely to have accidents, service failures, and be more likely to remain unemployed rather than turn over rapidly.

In fact, a pilot study (conducted by Personnel Decisions, Incorporated, publisher of the REI) investigating the potential utility of the REI, examined the relationship between REI scale scores and a global, subjective, supervisory rating of drivers' OTR performance proficiency. This study yielded a correlation of .22 between REI scores and ratings (p < .05) and a utility analysis estimated savings of 370,000 dollars per year in reduced turnover alone (see Appendix A).

The measurement of attitudes about responsibility, dependability, and conscientiousness, in combination with the promising results of the pilot study, led this OTR carrier to believe that the REI might be helpful to their selection system. Prior to implementing the REI in their selection system, however, it was felt that the relationship between the REI scores and both subjective and objective
and more scientifically controlled study. For this reason, an uncontaminated, predictive, criterion-related validity study using the REI was undertaken. Several hypotheses regarding the relationships between REI scores and both subjective and objective performance data were made.

1) It was hypothesized that drivers with higher REI scale scores would have fewer OTR performance problems. Performance problems would include driving accidents, service failures involving late pick-ups and late deliveries, and claims filed by customers for damaged freight.

2) It was also hypothesized that drivers with lower REI score might generate work performance problems which would be reflected in poorer objective data and subjective performance ratings. These problems would include the aforementioned performance problems and also diminished mileage driven (e.g., not accepting all loads assigned), fewer total loads (e.g., not accepting assigned loads or taking excessive time-off), and fewer monthly bonuses. Additionally, it was thought that drivers with lower REI scores would also earn lower subjective supervisory performance ratings.

3) It was further hypothesized that significant and positive correlations between REI scores, subjective performance ratings, and objective performance data would both replicate the findings of the pilot study and lend
support to the general hypothesis that the REI is predictive of OTR proficiency.

4) Also, it was hypothesized that drivers with higher subjective performance ratings would in fact have fewer performance problems and increased mileage, loads, and bonuses. In other words, drivers with good objective data scores would earn higher ratings from their supervisors.

5) Finally, it was hypothesized that the mean Tenure Score of the active sample would be significantly different and larger than that of the group of drivers who voluntarily or involuntarily terminated prior to the end of the study.

METHOD

Subjects

The subject pool was made up of applicants for the position of OTR driver at a large national trucking operation in Joplin, Missouri. In general, OTR applicants either have previous OTR driving experience or have attended a truck driving school. In this sense, the applicant pool shared a common characteristic, an interest in having an OTR truck driving job. Most applicants were self referred for employment and were asked to complete an REI as part of their application procedure. A hiring decision was then made based upon the existing selection procedure (e.g., REI scores were neither known nor used in the hiring decision). During the period from November 1986, through late January, 1987, 231 predominantly male applicants
late January, 1987, 231 predominantly male applicants
completed REI's. Of this initial pool, 26 were never hired
due to initial problems in the selection procedure (e.g.,
did not meet Department of Transportation, DOT,
requirements, falsified application and/or references,
failed drug screen, etc.), 84 were hired but subsequently
terminated during the study (e.g., refusal to follow company
rules, excessive moving violations, involvement in a major
accident, failed drug screen, etc.), leaving an "active"
pool of 121 drivers. All subjects were instructed that the
test they were taking would be used for research purposes
only and that their candor was thus encouraged. They were
also told that the test would in no way affect their chances
of employment.

With regard to demographics, subjects varied in terms
of age, sex, and race. The mean age in the sample was 34.2,
with a standard deviation of 4.5. Ninety-three percent
(93%) were male leaving seven percent female. Ninety percent
(90%) were caucasian, six percent (6%) Black, three percent
(3%) were American Indian, and one percent (1%) were
Hispanic.

Instrument

Subjects completed the Retail Employment Inventory
(REI) which is composed of 69 items requiring True/False
endorsements and 28 self-descriptive and background
multiple-choice questions. As discussed previously, the REI
is a psychometrically oriented pre-employment test which generates a performance and tenure score for each applicant.

The PDI Retail Employment Inventory was developed by George Paajanen and the staff of Personnel Decisions Incorporated (PDI) of Minneapolis, Minnesota and was first distributed for organizational use in 1985. With regard to underlying constructs, the REI is a short, pre-employment inventory designed to predict how trustworthy, dependable, responsible, and conscientious prospective employees are likely to be (Paajanen, personal communication, July, 1987). Theoretically, then, the REI should be of use to employers particularly interested in selecting applicants who are dependable and responsible.

In terms of its construction, the REI is composed of four scales which generate two scores: a performance score and a tenure score. The four scales are: performance predictor, tenure predictor, response frankness (an inverse lie scale), and response infrequency (an index of probable random responding). The four REI scales generate two scores for each applicant: a tenure score and a performance score. The tenure score is a prediction of how likely the applicant will be to turnover within three months, while the performance score is an indicator of how likely the applicant will be to engage in counter-productive job behavior.

In terms of its reliability and validity, the REI was validated in a predictive, criterion-related study
Personnel Decisions, Inc., 1985) in a large national retail chain. In its validation, a total of 4,652 applicants were tested and 2,988 were subsequently hired. Over 45 different hourly job titles were represented in the study, but the majority worked in receiving, stocking, and cashiering jobs. The employment status of those hired was monitored to determine how long they remained employed and if they were terminated what the reasons for the terminations were. The test-retest reliabilities of the scales were: performance predictor, .62; tenure predictor, .59; response predictor, .67; and response infrequency, .12. The results of this validation yielded validity coefficients between .25 and .35 (p < .01) for the performance and tenure scales with the related job performance criteria.

Procedure

The primary method employed in this study was allowing the subject pool sufficient time (3 to 6 months) to exhibit job related behavior. The design of this study, in its simplest sense, consisted of: a) having all subjects complete an REI; b) allowing their individual employment situations to develop (e.g., successfully or unsuccessfully completing their application process and exhibiting their OTR proficiency or lack thereof); c) collecting and recording these objective behaviors; d) scoring the REI's; e) gathering subjective performance ratings made by each subject's supervisor; and f) examining the statistical relationships between these elements.
After approximately 225 REI's had been completed, they were put aside for four months while the subject pool accumulated behavior ratings. During this waiting phase, it was the task of the investigator to: a) identify and select the objective variables which would capture a driver's actual OTR proficiency, and b) gather subjective driver performance ratings from driver supervisors for the drivers in the study.

After the six month period, the personnel files of the drivers who had remained active (e.g., those who were hired and had not been terminated prior to the end of the six month period) were carefully examined and information relevant to the 11 objective criteria was gathered. The subjective performance ratings were also completed at this time (e.g., after driver managers had had the opportunity to perceive and develop an impression of a driver's OTR proficiency).

The REI's were then scored using a hand held computer-like scoring machine provided by Personnel Decisions Incorporated. The subject's answers to each REI question were keyed into the machine and ultimately a Performance score and Tenure score was calculated. These scores were then recorded on the cover of the REI and the objective performance data sheet and the subjective rating form corresponding to that subject were placed inside his or her REI.
Finally, each subject's data (objective data, subjective ratings, and REI scores) were entered into a large data matrix and formatted according to ASCII guidelines. Correlations were then computed between all variables. It should be noted that due to the experimental design, subject's objective performance data were divided by a variable called Days of Service (the total number of days a subject were employed during the study) in an effort to equalize all subjects' opportunities to exhibit behavior.

Additionally, the mean Tenure Scores and Performance Scores earned by the active sample and earned by those terminated during the study were compared.

Criterion Measures

In terms of gathering objective performance data, the trucking company's system for tracking drivers was examined and potential criterion variables were identified. The information regarding objective driver performance was available from essentially two scores: a) files containing drivers' DOT history and personnel information, and b) banks of computer information regarding driver pay, mileage, and truck and load assignments.

From these two sources, it was apparent that there was an abundance of potential criterion related variables available. The criterion variables were chosen for their "ability" to capture the true performance of a driver; in other words, how well the driver was doing his job. The
eleven variables are: a) number of preventable accidents while driving for the company (NPRE), b) number of non-preventable accidents while driving for the company (NNON), c) total cost of all accidents during the study in dollars (ACC$), d) number of late pick-ups (LPU), e) number of late deliveries (LDE), f) the total number of loads hauled during the study (TOTLD), g) the percent availability of each driver for work on a monthly basis (BONUS): (100% availability = being able to work 27 out of 30 days per month); h) total number of claims pertaining to damaged or missing freight (CLAIM), i) total cost in dollars of these claims (CLAIMS$), and j) the total cost in dollars of any and all workman’s compensation claims made during the study (WCMP$).

With regard to subjective performance ratings, a form (Appendix B) with questions regarding global driver performance was developed from the form used in the initial PDL pilot study. These questions tap driver managers’ perceptions about driver’s level of responsibility, dependability, and conscientiousness. The four questions revolve around issues of driver performance, trustworthiness with difficult load assignments, comparisons with other drivers, and a willingness to call in and communicate directly with supervisors.

RESULTS

The descriptive statistics on all variables are presented in Table 1. It should be noted that zero values
Table 1
Descriptive Statistics on All Variables

<table>
<thead>
<tr>
<th>Var</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDI PERF</td>
<td>53.14</td>
<td>6.69</td>
<td>37</td>
<td>76</td>
<td>121</td>
</tr>
<tr>
<td>PDI TEN</td>
<td>25.23</td>
<td>3.58</td>
<td>15</td>
<td>35</td>
<td>121</td>
</tr>
<tr>
<td>RATING 1</td>
<td>3.92</td>
<td>1.05</td>
<td>1</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>RATING 2</td>
<td>3.68</td>
<td>1.11</td>
<td>1</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>RATING 3</td>
<td>3.44</td>
<td>1.02</td>
<td>1</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>RATING 4</td>
<td>2.06</td>
<td>.38</td>
<td>1</td>
<td>4</td>
<td>90</td>
</tr>
<tr>
<td>RATING 5</td>
<td>4.67</td>
<td>.49</td>
<td>3</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>EXP</td>
<td>31.23</td>
<td>32.34</td>
<td>0</td>
<td>96</td>
<td>121</td>
</tr>
<tr>
<td>NPRE</td>
<td>.01</td>
<td>.005</td>
<td>0</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>NNON</td>
<td>.015</td>
<td>.009</td>
<td>0</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>ACCS</td>
<td>4.52</td>
<td>11.48</td>
<td>0</td>
<td>87</td>
<td>120</td>
</tr>
<tr>
<td>LPU</td>
<td>.001</td>
<td>.004</td>
<td>0</td>
<td>0</td>
<td>119</td>
</tr>
<tr>
<td>LVE</td>
<td>.011</td>
<td>.015</td>
<td>0</td>
<td>0</td>
<td>119</td>
</tr>
<tr>
<td>GRMLG</td>
<td>223.75</td>
<td>78.25</td>
<td>5</td>
<td>443</td>
<td>111</td>
</tr>
<tr>
<td>TOTLD</td>
<td>.24</td>
<td>.098</td>
<td>0</td>
<td>0</td>
<td>98</td>
</tr>
<tr>
<td>BONUS</td>
<td>.008</td>
<td>.009</td>
<td>0</td>
<td>0</td>
<td>119</td>
</tr>
<tr>
<td>CLAIM</td>
<td>.000</td>
<td>.003</td>
<td>0</td>
<td>0</td>
<td>118</td>
</tr>
<tr>
<td>CLAIM $</td>
<td>.081</td>
<td>.730</td>
<td>0</td>
<td>0</td>
<td>118</td>
</tr>
<tr>
<td>WCMP $</td>
<td>2.821</td>
<td>10.037</td>
<td>0</td>
<td>58</td>
<td>118</td>
</tr>
<tr>
<td>DSV</td>
<td>86.512</td>
<td>39.219</td>
<td>22</td>
<td>168</td>
<td>121</td>
</tr>
</tbody>
</table>

in the Maximum column are in fact decimals greater than zero which are rounded to zero by a computer subroutine. These
variables have been divided by the number of days worked by each subject (days of service - DSV) in an effort to equalize all subjects' opportunity to exhibit behavior. It is these resulting decimals, then, that have been rounded to zero in this Table only (e.g., actual decimal values were used in the computation of correlations).

First, the findings of the present study fail to replicate the findings of the PDI pilot study. The reader should recall that this pilot study produced a significant and directionally appropriate correlation coefficient of .22 between REI scores and subjective ratings of drivers OTR proficiency.

The relatively weak and non-significant correlation coefficients computed between drivers' REI scores and subjective performance ratings within the present study are presented in Table 2.

Second, the majority of the correlation coefficients computed between REI scores and objective performance data were weak and non-significant. Significant correlations were found between: Tenure Score and Number of Preventable Accidents, $r = .20$, $p < .05$; and Tenure Score and Total Accident Costs, $r = .15$, $p < .05$; Tenure Score and Total Mileage, $r = -.19$, $p < .05$; Tenure Score and Total Number of Loads Hauled, $r = -.18$, $p < .05$; and Tenure Score and Percent Availability to Work (Bonus), $r = -.17$, $p < .05$. Unfortunately, these correlations, although significant, were not directionally appropriate in terms of pre-experiment hypotheses. The correlation coefficients computed between
Table 2

Correlation Coefficients Computed Between REI Scores and Subjective Performance Ratings (see Appendix C).

<table>
<thead>
<tr>
<th>Performance Ratinga</th>
<th>REI</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance Score</td>
<td>Tenure Score</td>
</tr>
<tr>
<td>Rating 1 (Reliability)</td>
<td>.10</td>
<td>.16</td>
</tr>
<tr>
<td>Rating 2 (Dependability)</td>
<td>.09</td>
<td>.19</td>
</tr>
<tr>
<td>Rating 3 (Overall rating)</td>
<td>.13</td>
<td>.10</td>
</tr>
<tr>
<td>Rating 4 (Conscientiousness)</td>
<td>.07</td>
<td>.28</td>
</tr>
</tbody>
</table>

Note. N = 89.

aRating 5 "is not" included because it is not a performance measure.

REI scores and objective performance data are presented in Table 3.

Third, correlation coefficients were computed between the 11 objective performance variables and the four subjective performance ratings. Significant and directionally appropriate correlations were found between several of the objective and subjective variables, and all of the coefficients are presented in Table 4.
Table 3

Correlation Coefficients Computed Between REI Scores and Objective Performance Data

<table>
<thead>
<tr>
<th></th>
<th>REI Performance Score</th>
<th>REI Tenure Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>NFRE ($n = 121$)</td>
<td>-.03</td>
<td>.34</td>
</tr>
<tr>
<td>NNON ($n = 121$)</td>
<td>-.07</td>
<td>.21</td>
</tr>
<tr>
<td>ACC$ (n = 120)$</td>
<td>-.06</td>
<td>.23</td>
</tr>
<tr>
<td>LPU ($n = 119$)</td>
<td>-.14</td>
<td>.06</td>
</tr>
<tr>
<td>LDE ($n = 119$)</td>
<td>.02</td>
<td>.39</td>
</tr>
<tr>
<td>GRMLG ($n = 111$)</td>
<td>-.12</td>
<td>.10</td>
</tr>
<tr>
<td>TOTLD ($n = 98$)</td>
<td>.07</td>
<td>.25</td>
</tr>
<tr>
<td>BONUS ($n = 119$)</td>
<td>.05</td>
<td>.27</td>
</tr>
<tr>
<td>CLAIM ($n = 118$)</td>
<td>.04</td>
<td>.32</td>
</tr>
<tr>
<td>CLAIM$ (n = 118$)</td>
<td>.10</td>
<td>.12</td>
</tr>
<tr>
<td>WCOMP $ (n = 118)$</td>
<td>-.04</td>
<td>.33</td>
</tr>
</tbody>
</table>

With regard to intercorrelations among the objective criterion variables, found in Table 5, several significant (33%) and directionally appropriate correlations are evident. Without listing each significant correlation, the investigator would like to point out that intuitively appropriate relationships between objective variables, for example the correlation of $\rho = .57, p < .01$, between the number of non-preventable accidents (NNON) and the total cost of accidents (ACC$), are borne out by the data.
Table 4

Correlation Coefficients Computed Between Drivers' Subjective Performance Ratings and Objective Performance Data

<table>
<thead>
<tr>
<th></th>
<th>Score 1 (Reliability)</th>
<th>Score 2 (Dependability)</th>
<th>Score 3 (Overall Proficiency)</th>
<th>Score 4 (Conscientiousness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPRED</td>
<td>-.13</td>
<td>-.09</td>
<td>-.09</td>
<td>-.04</td>
</tr>
<tr>
<td>(Preventable Accidents)</td>
<td>.10</td>
<td>.19</td>
<td>.19</td>
<td>.36</td>
</tr>
<tr>
<td>NNON</td>
<td>.01</td>
<td>-.02</td>
<td>-.06</td>
<td>-.05</td>
</tr>
<tr>
<td>(Non-Preventable Accidents)</td>
<td>.45</td>
<td>.41</td>
<td>.27</td>
<td>.32</td>
</tr>
<tr>
<td>ACC $</td>
<td>-.10</td>
<td>-.15</td>
<td>-.19</td>
<td>-.01</td>
</tr>
<tr>
<td>(Accident Costs)</td>
<td>.17</td>
<td>.08</td>
<td>.04</td>
<td>.46</td>
</tr>
<tr>
<td>LPU</td>
<td>-.12</td>
<td>-.14</td>
<td>-.15</td>
<td>-.04</td>
</tr>
<tr>
<td>(Late Pick-Ups)</td>
<td>.14</td>
<td>.09</td>
<td>.08</td>
<td>.35</td>
</tr>
<tr>
<td>LDE</td>
<td>-.18</td>
<td>-.27</td>
<td>-.24</td>
<td>-.11</td>
</tr>
<tr>
<td>(Late Deliveries)</td>
<td>.046</td>
<td>.005</td>
<td>.01</td>
<td>.16</td>
</tr>
<tr>
<td>GRMLG</td>
<td>.35</td>
<td>.29</td>
<td>.28</td>
<td>.09</td>
</tr>
<tr>
<td>(Total Mileage)</td>
<td>.001</td>
<td>.004</td>
<td>.005</td>
<td>.21</td>
</tr>
<tr>
<td>TOTLD</td>
<td>.28</td>
<td>.28</td>
<td>.29</td>
<td>.03</td>
</tr>
<tr>
<td>(Total Loads Hauled)</td>
<td>.008</td>
<td>.067</td>
<td>.006</td>
<td>.40</td>
</tr>
<tr>
<td>BONUS</td>
<td>.29</td>
<td>.31</td>
<td>.34</td>
<td>.02</td>
</tr>
<tr>
<td>(Willingness to Work)</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.41</td>
</tr>
<tr>
<td>CLAIM</td>
<td>.05</td>
<td>.04</td>
<td>-.02</td>
<td>-.02</td>
</tr>
<tr>
<td>(Customer Claims)</td>
<td>.32</td>
<td>.35</td>
<td>.42</td>
<td>.42</td>
</tr>
<tr>
<td>CLAIMS $</td>
<td>.11</td>
<td>.03</td>
<td>.05</td>
<td>-.02</td>
</tr>
<tr>
<td>(Cost of Claims)</td>
<td>.15</td>
<td>.37</td>
<td>.32</td>
<td>.43</td>
</tr>
<tr>
<td>WCMP$</td>
<td>.08</td>
<td>.14</td>
<td>.08</td>
<td>-.08</td>
</tr>
<tr>
<td>(Workman's Compensation Costs)</td>
<td>.22</td>
<td>.09</td>
<td>.22</td>
<td>.42</td>
</tr>
</tbody>
</table>
Table 5

**Intercorrelations between Objective Variables (N = 121)**

<table>
<thead>
<tr>
<th></th>
<th>NPRE</th>
<th>NNON</th>
<th>ACC$</th>
<th>LPU</th>
<th>LDE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NYPR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NNON</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.16</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACC$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>0.57</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LPU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.29</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.46</td>
<td>0.00</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td><strong>LDE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.39</td>
<td>0.46</td>
<td>0.37</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>GRMLG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.08</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.17</td>
<td>0.45</td>
<td>0.33</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>TOTLD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.06</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.38</td>
<td>0.37</td>
<td>0.38</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>BONUS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.04</td>
<td>-0.14</td>
<td>-0.17</td>
<td>-0.03</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.34</td>
<td>0.06</td>
<td>0.03</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>CLAIM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.25</td>
<td>0.12</td>
<td>0.27</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.35</td>
<td>0.01</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>CLAIM$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>0.07</td>
<td>0.05</td>
<td>-0.01</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.39</td>
<td>0.22</td>
<td>0.28</td>
<td>0.49</td>
</tr>
<tr>
<td><strong>WCNO$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.40</td>
<td>0.29</td>
<td>0.39</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5--Continued

<table>
<thead>
<tr>
<th></th>
<th>GRMLG</th>
<th>TOTLD</th>
<th>BONUS</th>
<th>CLAIM</th>
<th>CLAIM$</th>
<th>WCMPS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRMLG</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTLD</td>
<td>.64</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BONUS</td>
<td>.54</td>
<td>.65</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLAIM</td>
<td>-.05</td>
<td>.01</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLAIM$</td>
<td>.04</td>
<td>.18</td>
<td>.12</td>
<td>.44</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCMPS$</td>
<td>-.07</td>
<td>-.02</td>
<td>-.01</td>
<td>-.04</td>
<td>-.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

In a similar vein, the intercorrelations between the subjective performance questions correlate significantly as seen in Table 6.

The intercorrelations between REI Scale scores indicate that scores on one scale are indeed indicative of scores on the other scale. Performance Scores and Tenure Scores correlate $r = .31$, $p < .01$.

Finally, mean Tenure Scores and Performance Scores of the active sample and of the terminated sample were compared. The mean Tenure Score earned by the active subjects was
Table 6

<table>
<thead>
<tr>
<th></th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score 1 r</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score 2 r</td>
<td>.88</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score 3 r</td>
<td>.83</td>
<td>.86</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>Score 4 r</td>
<td>.24</td>
<td>.20</td>
<td>.17</td>
<td>1.00</td>
</tr>
<tr>
<td>p</td>
<td>.01</td>
<td>.03</td>
<td>.05</td>
<td>.00</td>
</tr>
</tbody>
</table>

25.2, and the mean Tenure Score earned by those subjects who voluntarily or involuntarily terminated during the study was 24.8. The mean Performance score earned by the active subjects was 53.14, and the mean Performance score earned by the terminated group was 52.23. These means were not significantly different when subjected to a t test.

Discussion

Since its inception, the design, procedure, and analysis within this investigation have been focused on determining the predictive validity of the REI in an OTR setting. With the results and analyses in hand, it is apparent that the present investigation indicates that REI scores are not highly predictive of OTR proficiency. This conclusion is
borne out by the experimental results in several areas.

First, the present investigation failed to replicate a pilot study finding which indicated that there was a significant and positive correlation between drivers' REI scores and subjective supervisory performance ratings. A driver's REI scores, then, do not seem to be related to how a driver's performance is perceived by his supervisor. The significant and directionally appropriate correlations found between subjective performance ratings and objective performance criteria indicate that driver managers do accurately perceive OTR proficiency; and, that the REI scores do not reflect OTR proficiency.

Second, the failure to find strong, directionally appropriate correlations between REI scores and objective performance variables indicates that within this study, the REI is not highly predictive of objectively measured OTR proficiency. While no significant correlations were found between the 11 objective variables and the REI's Performance Score, some significant correlations, although directionally inappropriate, were found between the Tenure Score and some objective criteria. These significant correlations include Tenure Score with Number of Preventable Accidents, $r = .20$, $p < .05$; with Total Accident Costs, $r = .15$, $p < .05$; with Total Mileage Driven, $r = -.19$, $p < .05$; with Total Number of Loads Hauled, $r = -.18$, $p < .05$; and with a monthly measure of willingness to work (BONUS), $r = -.16$, $p < .05$. 
The two correlations related to accident criteria appear important in terms of organizational expenses; however, the effects of sampling error (only seven drivers had preventable accidents) shade these findings to some degree. Designing a selection scheme around the remaining correlations would probably not save the organization money in terms of accident and turnover costs; however some gain might be made in terms of improving the driver force.

Finally, the efficiency of the Tenure Scale in terms of predicting tenure was evaluated by comparing the mean tenure scores earned by the subjects in the active group and in the terminated (both voluntary and involuntary) group. At t test revealed no significant difference between the group that remained employed and the group that did not in terms of tenure scores earned. This led to conclusion that the Tenure Scale was not discriminating between individuals who remained or did not remain employed.

These three findings, then, indicate that the REI is not highly predictive of OTR proficiency. This conclusion, although disappointing from a selection research standpoint, is valuable from an organizational perspective. Had the REI been included in the OTR company's selection strategy solely on the basis of the pilot study findings, tremendous costs to the organization may have been encountered on two fronts.
First, at a cost of approximately 10 dollars (\$10.00) per test, significant expense would have been incurred for materials alone. Second, and perhaps more importantly, the effect of using an inappropriate (not valid) test to screen applicants might be the exclusion of candidates who may in fact be well qualified for an OTR position. And third, the use of a non-valid selection test, the potential for adverse impact, and the possible exclusion of otherwise qualified applicants would be a clear violation of EEOC guidelines. For these reasons, then, the findings of the present study are both important and valuable.

These conclusions lead to the question of why the REI, a test designed to measure the likelihood of responsible, dependable, and conscientious job behavior, was not accurately identifying OTR drivers who were or were not performing their jobs proficiently. In an effort to clarify and answer this question, a confirmatory job analysis of the OTR position was completed through interviews with Subject Matter Experts (SME's). The goal of this job analysis, then, was to identify the personal characteristics which seem to separate good OTR drivers from bad OTR drivers. The rationale for completing this job analysis, although after the fact in timing, was the accurate identification of OTR proficiency-related characteristics which should be the basis for pre-employment screening.
Over the course of several days of hour-long interviews with about 30 SME's (OTR drivers, driver managers, operations people, and OTR Management), a brief job analysis of the OTR driver position was completed. In addition to discussion about the OTR position, each interviewee was asked to circle the five characteristics he or she felt were most important to OTR proficiency from a list of several characteristics (see Appendix C).

Upon completion of the job analysis, a frequency count of the characteristics (the number of times a characteristic was circled across all interviews) was performed and the results appear in Appendix C).

The clear emergence of the characteristics of Responsibility and Dependability confirmed their importance to OTR proficiency; however, these results merit hypotheses as to why the REI was not predictive of OTR proficiency. There are several possible explanations as to why REI scores did not correlate with subjective or objective driver performance data; however, three are particularly interesting.

First, issues related to the subject pool cannot be ignored. With regard to timing, this study was conducted during the annual slow period or slump in freight volume (July = High Freight Volume, December = Low freight Volume) (A.D. Thompson, personal communication, October, 1987). It is possible, then, that the applicants circulating in the OTR
environment during this time period are somewhat different than those circulating during better market conditions.

Second, and related to the characteristics of the subject pool, is the issue of test sensitivity. The types of responsible, dependable, and conscientious attitudes needed for OTR proficiency may be different from those measured by the REI. The initial validation of the REI was performed on retail employees, a job very different from that of the OTR driver. In fact, the intercorrelation achieved between REI scores were lower (.34 vs. .45) than those achieved in other REI validations (B. Sevy, personal communication, June, 1987) indicating that test sensitivity may have been a problem in this validation.

Third, the types of responsibility, dependability, and conscientiousness required for OTR proficiency must certainly be viewed in relation to the OTR context. The work of an OTR driver is very demanding for three reasons: a) OTR drivers work away from home for an average of 18 to 21 days at a time (A.D. Thompson, personal communication, July, 1987); b) they are responsible for an expensive truck, trailer, and whatever cargo they are hauling; and c) they are frequently exposed to stress and fatigue from driving, irregular sleep patterns, and from loading and unloading their trucks. Designing a test to select individuals who possess the types of responsibility and dependability needed
to succeed under these difficult conditions represents a significant psychometric challenge.

In terms of construct validity, the results of the present study appear to indicate that the types of responsibility and dependability needed for OTR proficiency are not captured within the constructs of the REI. In other words, the REI may not be tapping the OTR relevant variance within the broad dependability and responsibility constructs. Further research on this issue is indicated.

With regard to directions for future research, it is the opinion of the author that the reduction of turnover and unsafe driving behavior might be better controlled through programs designed to help drivers who are already employed rather than screening applicants prior to employment. As indicated earlier, designing and validating a test which would select individuals capable of surviving and excelling in the unique OTR job environment appears to be a difficult challenge. From this standpoint, it seems that taking steps to resolve retention issues within an intact OTR driver pool would be appropriate. Such activities as training to improve interpersonal skills for both drivers and driver managers, training in stress management, quality of work-life (QWL) programs for drivers, and the commitment to resolve job problem issues might help to alleviate increasing accident and turnover rates.
In conclusion, it seems that further research is needed in the area of OTR selection. While it is clear that tests need to be developed for the improved selection of truck drivers, larger gains in the areas of improved safety and lower turnover might be made from within the structure of the OTR operation. In light of the difficult and complex nature of this problem, easy and quick solutions are not available. Certainly, encouraging human beings to remain employed and behave safely is important to organizations for both financial and ethical reasons; and, hopefully, the efforts of I/O Psychologists can help committed organizations achieve this important goal.
Mr. 

Dear 

It was good talking with you the other day and hearing about your plans to increase the effectiveness of your driver selection decisions. I truly believe our Employment Inventory can be of great help in advancing this goal.

Enclosed are the materials and equipment you requested to evaluate our test. Please read the Administrator's manual carefully before using the test.

As we discussed, I want you to feel confident that the test will help you before you pay for it. However, in order to properly study the test's effectiveness, I will need your help. Specifically, you should have roughly 100 current drivers take the test. These drivers should be a fairly random sample, i.e., a mix of reliable, moderately reliable, and unreliable drivers. It might be a good idea to have each driver who stops in at ____ for a new load, maintenance, to pick up a paycheck, etc., take the test until 100 have taken it.

For each of these hundred drivers, we will need some data regarding their reliability, e.g., number of accidents, late deliveries, complaints, and any other measure which is available and important. I recommend creating a list with each driver's name, social security number and number of accidents, complaints, etc. indicated. You might also have each driver's supervisor rate the driver's reliability along a five point scale:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>highly reliable, one of the best</td>
</tr>
<tr>
<td>4</td>
<td>usually reliable, above average</td>
</tr>
<tr>
<td>3</td>
<td>fairly reliable, about average</td>
</tr>
<tr>
<td>2</td>
<td>somewhat unreliable, below average</td>
</tr>
<tr>
<td>1</td>
<td>very unreliable, one of the worst</td>
</tr>
</tbody>
</table>
We can then compare this data to each driver's test score to see how well the test identifies the most reliable drivers. This will do two things. First, it will help you immensely in the event of any challenge. Rather than simply using your opinion of the drivers as a criterion, you can say that the test identified drivers with less accidents, complaints against them, etc. Second, it helps us to gather further powerful, objective evidence of the test's validity.

It would be very helpful if you could record and send to us, along with each driver's scores, their number of accidents, complaints, late delivers, etc. within, say, the last year or six months.

I am excited about seeing the results and having you use our test to improve your driver force. We can send you cartridges to begin testing applicants within a day or two after receiving the results.

Please call me if you have any questions or if I can be of any further help.

Regards,

enclosures
Dear [Blank]:

Enclosed is a summary of the study we conducted to evaluate the effectiveness of our Employment Inventory for selecting drivers in your organization.

As you will see from this summary, [Blank] is in a position to realize a significant benefit by implementing the Employment Inventory. Based on the most conservative estimates, [Blank] can expect a payoff of over $300,000 over and above the $20,000 projected cost to screen 2,000 applicants.

It is important to note that these figures do not include cost savings due to screening out applicants early in the selection process, thereby reducing the number who undergo more expensive selection instruments. In addition, the Employment Inventory may prove to be effective in reducing turnover, since it has proven to be an effective device for predicting tenure in other organizations. To examine this, we can retain the test scores on the drivers who completed the Employment Inventory and follow up six or nine months from now to examine the relationship between test scores and turnover at [Blank].

Thank you, [Blank] for your efforts and support throughout this study. Please call me if you have any questions or if I can be of any further assistance.

Cordially,

[Blank]

cc: [Blank]
EVALUATION SUMMARY

Evaluation Objectives

To determine:

(1) The relationship between drivers' performance on PDI's Employment Inventory and various measures of driver job performance.

(2) The potential cost/benefit of using the Employment Inventory as a component in driver selection process.

Study Design/Process

(1) Administered the Employment Inventory to 96 current drivers.

(2) Collected information on each of the 96 drivers' job performance, including:
   - supervisors' performance ratings
   - number of accidents
   - number of claims for damaged merchandise
   - number of service failures

(3) Compared driver performance (item 2 above) to Employment Inventory scores to determine how the EI relates to each factor.

(4) Estimated the expected financial impact (cost or benefit) of using the Employment Inventory in current driver selection process.

Employment Inventory Effectiveness

(1) The EI related positively with driver reliability, as rated by drivers' supervisors. Drivers with higher scores on the EI tended to be rated as being more reliable and dependable.

(2) The EI correlated positively with the number of accidents. Drivers who scored higher on the EI tended to have more accidents than those who scores were lower.

(3) No relationship (either positive or negative) was found between EI scores and service failures or claims.

Several factors may account for these findings. First, damaged merchandise claims may be, in large part, beyond the control of the driver, since the way in which merchandise is loaded or unloaded may cause damage. To the extent that claims are beyond drivers' control, we would not expect a relationship between driver reliability and
Employment Inventory Effectiveness (continued)

number of claims. Second, it appears that "hot" loads tend to be assigned to more reliable drivers, rather than unreliable drivers. Since the more reliable drivers are assigned more of such loads and are working under tighter deadlines, they have more opportunity to accumulate service failures. As a result, we might expect that highly reliable drivers would have as many or more service failures than highly unreliable drivers. Third, it appears that drivers' supervisors avoid sending unreliable drivers into heavy traffic areas (e.g., major cities in the Northeast section of the country) or into rough weather (e.g., the Rocky Mountains during winter storms). Instead, the more reliable drivers tend to run those routes, giving them more exposure to accident risk. To the extent that this is true, the EI's positive relationship with accidents may simply be due to the fact that more reliable drivers are more likely to be assigned loads in areas with a high potential for accidents.

In summary, the variable over which drivers seem to have the most control, supervisors' ratings of their reliability, shows a significantly positive relationship to scores on the EI. While scores on the EI were not positively related to other measures of driver effectiveness, some questions remain regarding the appropriateness of these measures as valid indicators. It seems that both service failures and accidents could be more a function of the loads to which the driver is assigned than the reliability or dependability of the driver. Also, there is some question regarding the degree of control a driver has over damage to his/her load.

Financial Payoff/Return on Investment

Based on our findings, we would expect that implementing the EI would result in an increase in the effectiveness of S driver force. There is some question regarding whether selecting more effective drivers would actually increase the number of accidents or whether the positive relationship between driver effectiveness and number of accidents would simply disappear, since there would be less unreliable drivers to avoid sending into high accident risk regions. Given this uncertainty, we conducted a conservative analysis, assuming that implementing the EI would actually increase the number of accidents.

In conducting our analysis of the financial impact of implementing the EI, we made several assumptions:

1. That more effective driver performance results in a financial benefit for S

2. That ineffective performance results in a financial cost for S

3. That S hires 25% of all applicants for driver positions.

4. That S processes approximately 2,000 driver applications per year.
Financial Payoff/Return on Investment (continued)

Based on these assumptions, we determined the expected payoff from increased driver effectiveness as well as the expected cost of an increase in the number of accidents. The per driver payoffs and costs are shown below:

Financial payoff from increased driver effectiveness: $960/driver
Financial cost of increased number of accidents: $180/driver
PAYOFF $780/driver

While this payoff seems rather modest, it certainly adds up when 500 drivers are hired each year.

$780/driver \times 500 \text{ drivers} = $390,000

Subtracting the cost to administer the EI to 2,000 applicants ($20,000) leaves a net payoff of $370,000.

The actual financial payoff will vary, depending on the point at which the pass/fail or "cutoff" score is established. Based on an overall selection ratio of 25% (only one in four applicants is hired) and assuming that the EI would be used as a screening device early on the selection process, we would recommend that the EI be used to screen out 40%-60% of the applicants. Given this range, we estimate that can expect the following:

<table>
<thead>
<tr>
<th>% of Applicants Screened out</th>
<th>Payoff per driver</th>
<th>Total Payoff per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>$955</td>
<td>$477,500</td>
</tr>
<tr>
<td>50</td>
<td>$922</td>
<td>$461,000</td>
</tr>
<tr>
<td>40</td>
<td>$790</td>
<td>$395,000</td>
</tr>
</tbody>
</table>

In addition to the above payoffs, we would expect some payoff due to the increased efficiency of screening out more applicants early in the selection process, rather than sending a large number of applicants through more expensive screening devices later in the process. Also, the EI has been shown to relate positively to tenure in past studies and, depending on the results of the tenure study currently underway at additional payoffs due to decreased turnover may result.
Driver Selection Study

Driver Performance Rating Form

This form is intended to help a driver manager think about and record his/her perceptions of a driver's performance. When completing this form, please think only about the driver whose name appears at the top of this form. Also, please be as precise as possible - the results of these evaluations are confidential and will be used for research purposes only. Answer the following questions by circling the number next to the statement which best fits the driver named above.

1. How would you rate this driver's performance?
   5 Highly reliable
   4 Usually reliable
   3 Fairly reliable
   2 Somewhat unreliable
   1 Very unreliable

2. In the past, how willing have you been to assign this driver to a "hot" load (e.g. special time, mileage, or customer considerations)?
   5 Highly willing, the driver will surely succeed
   4 Quite willing, driver will almost certainly succeed
   3 Somewhat willing, driver will probably do O.K.
   2 Reluctant, the driver might not do well
   1 Highly reluctant, the driver will almost surely fail

3. How would you rate this driver compared to other drivers?
   5 One of the best
   4 Above average
   3 About average
   2 Below average
   1 One of the worst

4. On average, how many times in a day do you have contact with this driver, either by telephone, phone mail, or in person?
   1 Less than once
   2 One to Three times
   3 Four to six times
   4 Seven to nine times
   5 Ten times or more per day

5. How confident are you in the accuracy of your ratings?
   5 Completely confident
   4 Confident
   3 Somewhat confident
   2 Slightly confident
   1 Not confident
APPENDIX C

FREQUENCY COUNT OF JOB RELATED CHARACTERISTICS SELECTED BY SME'S AS BEING MOST IMPORTANT TO OTR PROFICIENCY
Table 7

Frequency Count of Job Related Characteristics Selected by SME's as Being Most Important to OTR Proficiency

<table>
<thead>
<tr>
<th>CHARACTERISTIC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Law-Abiding</td>
<td>------</td>
</tr>
<tr>
<td>Persistent</td>
<td>------</td>
</tr>
<tr>
<td>Calmness</td>
<td>------</td>
</tr>
<tr>
<td>Poised</td>
<td>----</td>
</tr>
<tr>
<td>Energetic</td>
<td>----</td>
</tr>
<tr>
<td>Friendly</td>
<td>------</td>
</tr>
<tr>
<td>Self-Confident</td>
<td>------</td>
</tr>
<tr>
<td>Reasoning Ability</td>
<td>------</td>
</tr>
<tr>
<td>Motivated</td>
<td>------</td>
</tr>
<tr>
<td>Dependable</td>
<td>------</td>
</tr>
<tr>
<td>Patient</td>
<td>------</td>
</tr>
<tr>
<td>Tolerant</td>
<td>------</td>
</tr>
<tr>
<td>Alert</td>
<td>------</td>
</tr>
<tr>
<td>Responsible</td>
<td>------</td>
</tr>
<tr>
<td>Written Communication</td>
<td>------</td>
</tr>
<tr>
<td>Oral Communication</td>
<td>------</td>
</tr>
<tr>
<td>Emotionally Health</td>
<td>------</td>
</tr>
<tr>
<td>Independent</td>
<td>0_5_10_15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Somewhat Important</th>
<th>Important</th>
<th>Very Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>


REFERENCES


development based on a content oriented strategy.

*Personnel Psychology, 39, 91-108.*

