

379
N81
No. 6604

VALIDATION OF A TEST BATTERY FOR THE SELECTION OF
DRIVER MANAGERS IN A TRUCKING ORGANIZATION

THESIS

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

For the Degree of

MASTER OF SCIENCE

By

Shirley T. Cassel, B. A.

Denton, Texas

May, 1990

MB

Cassel, Shirley T. Validation of a Test Battery for the Selection of Driver Managers in a Trucking Organization. Master of Science (Industrial/Organizational Psychology). May, 1990, 78 pp., 5 tables, references, 55 titles.

This study was a concurrent validation of a paper and pencil test battery used at a national trucking company. Forty-eight driver managers were rated by their immediate supervisors with the performance appraisal covering 12 dimensions of job behavior that was developed by the experimenter. The driver managers were also administered the Wesman Personnel Classification Test, the Watson-Glaser Critical Thinking Appraisal, and the California Psychological Inventory (CPI). A biographical information blank was also developed and validated. Most validity correlations were nonsignificant, with the exception of the Dominance scale $r = .25$ ($p < .05$), the Self-control scale $r = -.25$ ($p < .05$), the Communality scale $r = .29$ ($p < .05$), and the Flexibility scale $r = -.39$ ($p < .05$), with overall performance.

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
ACKNOWLEDGEMENTS	v
VALIDATION OF A TEST BATTERY FOR THE SELECTION OF DRIVER MANAGERS IN A TRUCKING ORGANIZATION	
Introduction	1
Personnel Selection	
History of Paper and Pencil Testing	
Criterion Issues	
Design of Validation Research	
Statement of the Problem	
Statement of the Hypotheses	
Method	12
Subjects	
Design	
Measures	
Procedures	
Results	24
Discussion	35
APPENDICES	42
REFERENCES	72

LIST OF TABLES

Table	Page
1. Correlation of the Wesman PCT and the Watson-Glaser CTI with Ratings for the Total Carriers .	28
2. Correlation of the CPI with Ratings for All Carriers	29
3. Correlation of Biodata with Behavioral Ratings for Total Carriers	31
4. Varimax Rotated Maximum Likelihood Factor Loadings for the Criterion Measure	33
5. Factor Correlations with the Predictors	34

ACKNOWLEDGEMENTS

I would like to extend my appreciation to my family and friends for their love and support during this endeavor.

VALIDATION OF A TEST BATTERY FOR THE SELECTION OF
DRIVER MANAGERS IN A TRUCKING ORGANIZATION

Personnel selection is a process culminating in a decision to hire one or more applicants for employment and not to hire others (Guion & Gibson, 1988). It is of prime importance to select candidates that will be successful from the standpoint of both employee and employer. Spooner (1985) states that a better match of candidate capabilities to job requirements will enable companies to hire people who will stay on the job longer and will be more satisfied in their position.

A central issue in the area of personnel selection has been the development of various ways of showing that a selection procedure relates to successful performance on the job. Professionals agree that this can be accomplished by developing a validation strategy that describes the components of behavior effectiveness and having predictor measures that forecast effective job behaviors.

Inefficient methods of selection for management positions are prevalent in major trucking organizations. Increasingly, trucking companies are experiencing a critical shortage of management personnel and are continually dissatisfied with the unpredictable performance of new hires. Moskal (1989) reported that there is a

middle and upper management talent void in the rapidly growing truckload sector of the motor-carrier industry. Moskal also noted that the industry is not attracting or retaining new professional management personnel because of low profit margins and low compensation levels.

The present study addresses the need to become more efficient in the selection of key management positions by developing a validation strategy for the driver manager position in a national trucking company. Success in driver management positions has historically been difficult to predict. Therefore, a test battery that was used primarily for both selection and career development purposes for the driver manager position, will be validated to explore its relationship to generalized performance ratings.

History of Paper and Pencil Testing

Many companies today utilize paper and pencil tests for selection because of their relative ease of administration and low costs. The use of tests for personnel selection began over 75 years ago with Munsterberg's experiment with motormen (Munsterberg, 1913). Ghiselli (1973) noted that large scale testing of both soldiers and industrial workers during World War I sparked new methodology and enhanced respectability to the examination of the use of tests in the assessment of occupational aptitude. Selection with personnel tests

remained a primary concern of industrial psychologists throughout World War II.

According to Ghiselli (1973), there has been an accumulation of experience with the use of professionally developed tests as devices for assessing men and women for positions in business and industrial establishments for well over 60 years. However, the use of these tests has been subject to controversy for many years, stifled by government regulation, conflicting guidelines, and lengthy litigation (e.g., Kleiman & Faley, 1985). There has been much debate within the psychological profession regarding tests and other methods used in making selection decisions (Grant, 1980).

Despite the controversy, the use of paper and pencil tests for personnel selection is increasing. Taylor and Zimmerer (1988) implied that because judges and employers know considerably more about these instruments, organizations have resumed their use. Strengthening this stance, recent research has shown that cognitive ability is the best indicator of future job performance, better by far than education, experience, or even the personal interview in the absence of behavioral data (Anastasi, 1982).

Guion and Gibson (1988) noted that from research using validity generalization, it is accepted that cognitive tests are likely to be good predictors of job performance.

Lent et al. (1971) found that general mental abilities correlated with supervisors' ratings for managerial positions $r = .48$. Furthermore, Ghiselli (1973) found that tests for intellectual abilities tended to be the best predictors of job performance criteria for managers, with the average validity coefficient for managerial occupations being $r = .29$. In subsequent meta-analysis research, Schmidt and Hunter (1981) and Hunter and Schmidt (1983) discovered that cognitive ability tests, as typically used in personnel selection, are valid predictors of work performance across a wide range of jobs.

Criterion Issues

As Cascio (1987) states, competent criterion research is one of the most pressing needs of personnel psychology today. Criteria are standards that can be used as yardsticks for measuring employees' success or failure (Bass & Barrett, 1981; Guion, 1965). There are many examples of possible criteria including: units produced, number of errors, number of days absent, length of service and ratings of behavioral expectations. These criteria aid in managerial selection as well as validation research.

Although objective indices of job behavior (e.g., salary, accidents, etc.) are intuitively appealing, these represent factors that are beyond the individual's control. As Cascio (1987) states, they do not measure behavior per

se, but rather outcomes of behavior. This is especially true for managers. There are many factors over and above the actions of the manager that affect the profit picture of his or her unit.

Subjective criteria, which include supervisor's ratings, are the most frequently used criteria (Guion, 1965; Lent et al., 1971). In a review of the criteria used for test validation, Guion (1965) indicates that superior's ratings of job behaviors are used about twice as frequently as more objective measures. Unfortunately, supervisors ratings are also susceptible to all forms of bias or rater errors.

Heneman (1975) defines performance appraisal as the "Achilles heel" of personnel management. Even though there are many disadvantages of performance appraisal, surveys of managers show that they are unwilling to abandon it because of its importance as an assessment tool (Labor Letter, 1984). Cascio (1987) concludes that since subjective measures depend upon human judgment, they are prone to certain kinds of errors associated with the rating process and that to be useful they must be based on careful analysis of the behaviors important for effective job performance.

One of the most common rater errors is "halo," which is a rater who assigns ratings on the basis of a global

impression of the ratee. Leniency and severity are also very common rater errors. This occurs when raters are inordinately lenient or inordinately difficult when rating the ratee. Central Tendency is another common rater error. This occurs when the rater avoids using the high and low extremes of the scales and tends to cluster all ratings about the center of all scales.

One way to control for these errors would include rater training, which ranges from a five minute review of the errors (Borman, 1975) to a three day seminar to control errors (Latham, 1975). Another method of controlling for errors is in the rating scale itself. There are many ways to format a scale to control for rater errors, but the most popular system, and the most acceptable to raters, is the behaviorally anchored rating scales (Smith & Kendall, 1963).

These behaviorally anchored rating scales use actual behaviors that are required to successfully function in a job. This allows the rater to relate incidents that the ratee has displayed on the job, when rating them. Tziner (1984) found that, compared with a graphic scale, a behaviorally anchored rating scale showed less leniency and less halo, as well as higher interrater reliability. According to Landy and Farr (1980), realizing the importance of performance measurement and actually measuring performance accurately are two different matters.

The careful development of criteria is an essential part of any validation study. Controlling rater errors and rating format can facilitate the rating process producing a more accurate appraisal.

Design of Validation Research

An early validation research paradigm was laid out by Freyd (1923) in a 10-step outline. Freyd established what has become the classical view of personnel research. Guion (1987) later abbreviated the outline:

- (1) analyze the job, (2) designate a single criterion,
- (3) select the subjects for the "experiment," (4) develop a list of abilities required for success (KSAP's), (5) find or devise a way to measure these abilities, (6) administer the measures "under carefully controlled conditions," (7) correlate test scores with the criterion, (8) combine measures for the best composite, (9) see if you could justify the new measures by comparing the results to those using the old methods of selection, and (10) if they offered an improvement, start using them. (p. 783)

Guion also noted that Freyd's techniques were marvelously up to date.

There are many ways to conduct validation research and utilizing the most appropriate method is an important issue. Guion (1976), along with many other industrial/

organizational psychologists, prefers criterion-related validation, especially the predictive model, for evaluating selection procedures. This method focuses on the measure of performance in the occupation. Cascio (1987) recommends criterion-related evidence of validity when measures of individual differences are used to predict behavior. When considering a criterion-related design, several issues should be considered, including the design, the predictors, and the criterion.

Cascio (1987) suggested that predictive validation designs were superior to concurrent designs. In contrast, Lent, Aurbach, and Levin (1971) found that the higher frequency of statistically significant results in predictive validity studies has not been due to design considerations but rather to more careful predictor and criterion development. Bemis (1968) found that with an empirical comparison of a large number of predictive and concurrent validities of the General Aptitude Test Battery (GATB), the two research designs yielded virtually identical validity coefficients.

In a subsequent study, Schmitt, Gooding, Noe, and Kirsch (1984) found that concurrent designs actually produce validity coefficients which are slightly superior to predictive designs. Also, Lent, Levin, and Aurbach (1971) found that concurrent validity studies offer an

advantage over predictive validity studies with respect to the measured relationship of the predictor and criterion. They pointed out that "the fact of near-simultaneity in time minimizes the attenuating influence of other factors upon the measured association between the experimental variables" (p. 527).

According to Barrett, Phillips, and Alexander (1981), there are several concerns with concurrent validity designs, including the missing person problem, restriction of range, differences between applicants and present employees, and the effects of job experience and training. The first criticism addressed by Barrett et al. (1981) was that concurrent validation studies "leave out" subjects that would be included in predictive validation designs. This is labeled the missing persons problem. They concluded that the missing person problem and restriction of range should be considered as one category. Range restriction for predictive validity designs was found to be very similar to range restrictions for concurrent validity designs. In both predictive validity and concurrent validity designs, not all available applicants would be hired, which would lead to a nonrandom sample. Fortunately, methods to correct for range restriction in criterion-related designs are available (see Lee, Miller, & Graham, 1982; and Linn, Harnisch, & Dunbar, 1981).

A second concern with concurrent validation methods involves the possibility that there are motivational differences between applicants and present employees that would affect scores on predictors. Barrett et al. (1981) found that the differences affected personality measures but not cognitive ability measures. They also found that test-taking motivation was correlated with work motivation which should cancel any motivational effects. Also, test-taking motivation may act in a purely random fashion, usually resulting in a spuriously low validity coefficient.

Furthermore, Barrett et al. (1981) reported that the effects of job tenure and training could affect scores on predictors and criteria. Therefore, if job experience and training were controlled, the learning would affect a job performance measure equally for predictive and concurrent studies. Barrett et al. (1981) concluded that for cognitive ability tests, there are no differences in results for predictive validity or criterion validity designs. In response to Barrett et al. (1981), Guion and Cranny (1982) argued that there are distinct differences between predictive and concurrent designs and one cannot simply substitute one procedure for another without considering the characteristics of a particular situation.

Statement of the Problem

In the organization under study, there were seemingly inefficient methods of selecting qualified personnel in the driver management position. Finding an appropriate device that discriminates between subsequently high and low performing employees was of utmost concern. The present study sought to address this question and hopefully facilitate the selection process for the driver management position.

Statement of the Hypotheses

Hypothesis one. No relationship will be revealed between behaviorally anchored, supervisory ratings of performance of driver managers and general intelligence as measured by the Wesman Personnel Classification Test.

Hypothesis two. No relationship will be found between behaviorally anchored, supervisory ratings of performance of driver managers and critical thinking as measured by the Watson-Glaser Critical Thinking Appraisal.

Hypothesis three. No relationship will be revealed between behaviorally anchored, supervisory ratings of performance of driver managers and personality dimensions as measured by the California Psychological Inventory.

Hypothesis four. No relationship will be found between behaviorally anchored supervisory ratings of performance of driver managers and responses to a biographical information form.

Method

Subjects

Subjects consisted of 48 driver managers, 46 of whom responded to the majority of the tests. Thirteen of the subjects were driver operator coordinators. These subjects have job duties that are similar to the driver managers. The driver managers worked for a major national trucking company in which they had direct supervisory control of "over-the-road" truck drivers. Thirty-nine of the 48 who participated were men and nine were women. The average age of a driver manager was 33, with the oldest being 49 and the youngest 18 years of age. The majority of the subjects were Caucasian, 92% were White, 4% were Black, and 4% were Hispanic.

Forty-four percent of the driver managers were high school educated, 39% had some college or technical training and 17% had a college degree. Most of the driver managers were married (76%), with 15% being single and 9% being divorced. Sixty-one percent of the driver managers had never driven a truck for a living. Seventy-eight percent revealed that they had experience in a "service" type position, while 33% had experience in the military.

Fifty-seven percent of the participants were smokers while 43% were nonsmokers. The majority of the driver managers owned their own home (65%), while 35% rented, and

driver managers drove an average of 16 miles to work. Most of the driver managers have been employed by the present company for an average of three years, with a standard deviation of two years. The average number of years a driver manager has been in the trucking industry is 7.65 years, with a standard deviation of seven years. The typical driver manager supervised an average of 44 drivers with a standard deviation of 13.16. The driver managers were responsible for an average of 36 trucks with a standard deviation of 3.86.

The subjects represented one of the largest national truckload carriers in the nation with 2,000 trucks in the combined fleet. Four cities with an average population of approximately 50,000 people were represented, including cities in the south, the northwest, the southeast, and the midwest. These operating companies were linked to a parent company which provided administrative support. All subjects were instructed that their participation was voluntary and that they would not be identified individually in this study.

Design

A concurrent validation design was considered the only practical option for this study because the tests were already in use. Also, historically there were few applicants and low turnover for the driver manager

position. Concurrent designs can facilitate the development of a performance appraisal system (Cascio, 1987) which was simultaneously being developed to serve as the criterion. The present study occurred in four phases. The first phase entailed generating a behaviorally based performance appraisal form (see Appendix A). Factors, and ultimately behavioral anchors were drawn from an extensive job analysis (see Appendix B). The ratings served as the criterion in the study. The second phase of the study involved training the performance raters on the rating process and collecting the criterion data. The third phase involved developing the biographical information form (see Appendix C). The fourth phase entailed administration of the test battery to be validated, to subjects who had not already taken it.

Selection Measures

Ryan and Sackett (1987) conducted a survey of individual assessment practices by industrial/organizational psychologists. It was found that the most frequently used ability tests were the Watson-Glaser Critical Thinking Appraisal (37.8%) and the Wesman Personnel Classification Test (19.3%). It was also found that 77% of I/O psychologists reported using personality inventories. The most commonly used instruments were the 16PF (33%), the Guilford-Zimmerman Temperment Survey (33%),

the CPI (28%), the Minnesota Multiphasic Personality Inventory (MMPI) (20%), and the Myers Briggs Type Indicator (19%). With this high rate of personality inventory usage, there seems to be surprisingly little new information on their usefulness, even though they continue to be used for selection of a wide range of jobs including executives (Wysocki, 1981) and groups such as police officers (Parisher, Rios, & Reilly, 1979).

Evidence indicates that the accuracy of biographical data as predictors of future work behavior is superb when appropriate procedures are followed (Cascio, 1987). It is also unlikely that biographical information will be falsified because this information can be verified. Lent et al. (1971) found that when personal information was used with supervisory evaluations, 24% of the time the correlations were statistically significant.

Wesman Personnel Classification Test. The present study utilized the Wesman Personnel Classification Test which is composed of 40 two-part analogy items arranged in increasing difficulty and 20 arithmetic computational items progressing from simple addition through more complex manipulations. The verbal analogies were the only section used in the present study. This paper-and-pencil test was being used by the company and had not been investigated as to whether or not it was useful in the selection of

employees. There are two forms of the test designed to measure general intelligence of a population from grades eight to 16 and adults. This test is primarily meant to be used in business settings. The norms are based on samples from production workers, executive trainees and high level positions as well as student groups. The manual (Wesman, 1985) reports reliability coefficients in the low 80's, and validity coefficients were reported as being higher for higher level jobs, including upper executive positions. This test is most suitable for white-collar employees, such as clerical, supervisory, and managerial personnel.

Research with the Wesman Personnel Classification Test (PCT) has produced many predictive validity studies (see Abt, 1949; Ash, 1960; Vincent & Dugan, 1962) which indicate that it has been found useful as a predictor of success in a wide variety of industrial applications. Furthermore, the Wesman PCT's validity studies report significant validity coefficients for a wide range of criteria of job success including overall and composite performance ratings. Finally, the PCT's split-half and test-retest reliability studies indicate coefficients in the low .80s which means that the scores should remain approximately the same after retesting.

Watson-Glaser Critical Thinking Appraisal. The present study also utilized the Watson-Glaser Critical

Thinking Appraisal (CTA) which is another general abilities test. This test was being used by the company, but had not been researched as to whether or not it was useful to the company as a selection tool. The Watson-Glaser CTA consists of 80 items which purport to measure five skills including (a) inference, (b) recognition of assumptions, (c) deduction, (d) interpretation, and (e) evaluation of arguments (Watson & Glaser, 1980). See Appendix D for scale definitions. One can use the subtests as an educational tool for remediation, but the manual stresses the use of the total score only (Watson & Glaser, 1980). There are two forms of the test and the reading level does not exceed ninth grade. National norms were established with high school students, college students, and business and civil service employees. The reliability coefficients (in the high .70s and low .80s) were reported in the manual and the only references of validity evidence pertained to content-related studies. Watson and Glaser (1980) suggest using the Watson-Glaser CTA when selecting candidates for positions where careful analytic reasoning is an important part of the job. Many validation studies have been conducted (i.e., Modjeski & Michael, 1983; Westbrook & Seelers, 1967) providing evidence of its usefulness in many diverse business and academic settings.

California Psychological Inventory. Personality tests are instruments for the measurement of emotional, motivational, interpersonal, and attitudinal characteristics as distinguished from abilities (Anastasi, 1982). Guion and Gibson (1988) noted that "from the perspective of personnel selection, personality may be more usefully narrowed to consistencies in behavior patterns relevant to the work to be done" (p. 352). With these concepts in mind, Guion and Gottier (1965) referred to a genuine need to predict behaviors influenced by personality-- the "will do" as opposed to the "can do" aspects of behavior on the job.

Along with the tests of general mental abilities, the new 1987 version (Gough, 1987) of the California Psychological Inventory (CPI) was used. The CPI was being used by the company and had not been researched as to whether or not it was useful in the selection of employees. The CPI is a self-administered, paper-and-pencil, personality test consisting of 462 items (e.g., "when I work on a committee I like to take charge of things."), 12 of which are duplicates. One hundred and seventy-eight of these items were taken directly from the Minnesota Multiphasic Personality Inventory (MMPI) item pool. All items are presented in a true-false format. There are 20 possible scores including: Dominance (Do), Capacity for

status (Cs), Sociability (So), Social presence (Sp), Self-acceptance (Sa), Independence (In), Empathy (Em), Responsibility (Re), Socialization (So), Self-control (Sc), Good impression (Gi), Communality (Cm), Well-being (Wb), Tolerance, (To), Achievement via conformance (Ac), Achievement via independence (Ai), Intellectual efficiency (Ie), Psychological mindedness (Py), Flexibility (Fx), and Femininity/Masculinity (F/M). For definitions of these scales, see Appendix E.

The CPI is designed for group administration but it can be taken individually or even by mail (Megargee, 1972). Standardized testing conditions are not essential and there is no time limit imposed. Fourth grade reading ability is required unless items are read aloud. The age range suggested for this test is 12 to 70 years of age.

An overall estimate of the test-retest reliability data for all CPI scales was made from the medians of each scale which resulted in a reliability median of .70 (Gough, 1987). The new CPI does not report validity coefficients, although the manual does report correlations of the CPI scales with three categories of tests and measures, including measures of intellectual and cognitive functioning, measures of aesthetic orientation or preference, measures of moral reasoning or moral development, personality and interest inventories, and assorted others.

The CPI, which has earned the respect of many academics, clinical psychologists, and human resource executives is an increasingly popular research tool. It has been one of the most extensively researched personality inventories (see Chronbach, 1959; Kelly, 1965; Thorndike, 1959), second only to the MMPI. The CPI has been used in business settings to predict academic achievement, leadership potential, and management success (see Bogard, 1960; Goodstein & Schrader, 1963; Hakstein, Woolsey, & Schroeder, 1987; Rawls & Rawls, 1968; Zdep, 1969).

Biographical Information Form. A biographical information form (see Appendix C) was developed to aid in describing the present driver management population. Some of the items included on the form were age, sex, educational level, and tenure.

Procedures

Development of a performance appraisal system was necessary because of the lack of comparable, objective, or subjective performance measures across the four companies. These scales were developed collaboratively by the experimenter and the company (see Appendix B) as a measure of job performance.

Phase one: Criterion development. A pre-existing job analysis, a concise job description, and subject-matter-experts (SME) were used in the development of the rating

form. The SMEs were top management employees who were familiar with the driver manager position and had worked with the driver managers or trained them. The job analysis consisted of statements of behavioral activities that occurred on the job, with comments concerning the appropriateness of the activity for the driver manager position. This information was compiled into dimensions by the researcher and presented to the experts for review. The experts were asked if the definitions were consistent with the dimensions, if the dimensions were redundant, and to suggest additional dimensions when needed. The dimensions were then revised and sent to the SMEs for final approval. Then a set of behavioral statements that would later become the behavioral anchors (Appendix E) were generated from the subject-matter-experts. They were asked to give examples of a behavior that was suitable for a score of "1" (Fails to meet performance expectations), "3" (Meets performance expectations) and "5" (Far exceed performance expectations). The list of behavioral examples provided by the SMEs were then compiled and returned to the SMEs for final approval. The behavioral anchors were placed on separate pages because of the large number of behaviors and to keep the basic rating form uncluttered.

The final 13-item rating scale consisted of items associated with nine basic dimensions of driver manager

effectiveness: driver relations, customer relations, sales and marketing relations, DOT (Department of Transportation) and safety compliance, personal work habits and stress management, general management of drivers, cost management, data accuracy, and timeliness. See Appendix B for complete definitions of criterion scales. Scale items were presented in a five point Likert-type format with responses ranging from 1 ("Fails to meet performance expectations") to 5 ("Far exceeds performance expectations"). Accuracy and timeliness scales were altered on the final scale because they were considered by the SMEs to be better represented with different anchors. Scale items for accuracy and timeliness were presented in a five-point Likert-type format ranging from 1 ("Below average performance"), to 3 ("Average performance") to 5 ("Standard performance"). The scale uses concise definitions of the dimensions being measured and a separate set of behavioral anchors which were provided to assist the rating process.

Phase two: Ratings. The driver managers' immediate supervisors, which included the Fleet Managers and the Vice President of Operations, received a general introduction to the research and a brief explanation of the rating scale used in the study. The driver manager's supervisors were instructed to read the handout while the investigator orally reviewed and discussed the information. The handout

(Appendix F) included a description of (a) common rater errors (i.e., leniency/severity), (b) warnings not to provide overall evaluations of a ratee by rating the individual at the same level on all dimensions, and (c) reminders that often job incumbents display strong points and weak points in their performance and that the ratings should reflect those strengths and weaknesses. The supervisors were informed that the purpose of the performance appraisal was for research purposes only (see Appendix G). The training session lasted approximately 20 minutes. Using the rating scale described earlier, the supervisors, who directly oversaw the driver managers activities and were all familiar with the driver manager job, independently evaluated their driver managers.

Phase three: Biodata development. A search for appropriate items for the background information form was found by asking subject-matter-experts what biographical information would be helpful to them. Some of the items were required for descriptive purposes and the rest of the items were developed from managerial suggestions and requests.

Phase four: Testing (Predictor data). The three paper-and pencil tests described earlier (the Wesman PCT, the Watson-Glaser CTA, and the California Psychological Inventory) were administered to the driver managers who had

not yet been tested by trained test administrators at the driver manager's location. These were then sent to the corporate headquarters for scoring. Some of the test scores utilized by this study were from previous testing of the driver managers. The driver managers were sent a cover letter (Appendix G) stating the purpose of the testing, that they would not be identified individually, and that they could decline to take part if they had strong objections.

Data Analysis

Pearson correlation coefficients were calculated between the scores obtained on the predictive tests and the subjective ratings of the driver managers to test the four hypotheses stated earlier. Also, a factor analysis was conducted of the items on the criterion measure to investigate the factor structure of the rating scale.

Results

The mean score for driver managers on the Wesman PCT verbal portion was 19.50 with a standard deviation of 6.88 and a range from six to 34. The driver managers' mean can be compared to norms for the Wesman. The closest category would be supervisors and with that normative group a score of 20 falls at the 39th percentile.

The mean for the Watson-Glaser CTA was 53.89 with a standard deviation of 10.20 and a range from 33 to 73. The

closest norm category would be sales representatives and with that normative group a score of 53 falls at the 25th percentile. See Appendix H for further statistics for the Wesman PCT and the Watson-Glaser CTA.

Appendix I shows the means and standard deviations of the California Psychological Inventory, in standard scores, of male and female driver managers. The CPI average (standardized) score of the CPI is 50, with a standard deviation of 10. Most of the driver managers' scores stayed within one standard deviation of the mean except on the Dominance scale. The Dominance scale for males (60.59) is 10.59 points above the mean, slightly more than one standard deviation. The females' mean on the Dominance scale is also somewhat high (57.44). Another point of interest in Appendix I includes the low scores on the Flexibility scale for both males and females. These two scores are the lowest scores for all of the CPI scales with males scoring 43.19 and females scoring 44.00.

Hypotheses Results

Four hypotheses were addressed in this study. The first hypothesis states that there is an absence of relationship between performance of driver managers on the Wesman Personnel Classification Test with behaviorally anchored supervisory ratings. According to Table 1, this hypothesis appears to be supported with the exception of

the correlation of the "availability" subscale of the criterion with the Wesman PCT ($\underline{r} = .30, \underline{p} < .05$).

The second hypothesis states that there is an absence of relationship between performance of driver managers on the Watson-Glaser Critical Thinking Inventory with behaviorally anchored supervisory ratings. According to Table 1, this hypothesis is supported with the exception of the "availability" subscale ($\underline{r} = .28, \underline{p} < .05$) and the "personal work habits and stress management" subscales, ($\underline{r} = .24, \underline{p} < .05$).

The third hypothesis states that there is an absence of relationship between performance of driver managers on the California Psychological Inventory with behaviorally anchored supervisory ratings. As shown in Table 2, there are several criterion subscales that correlated and the criterion total scale significantly correlates with the Dominance scale ($\underline{r} = .25, \underline{p} < .05$), Self-control scale, ($\underline{r} = -.25, \underline{p} < .05$), Communality scale ($\underline{r} = .29, \underline{p} < .05$), and Flexibility scale ($\underline{r} = -.37, \underline{p} < .05$). "Data Accuracy" and "Timeliness" were taken out of the scale because the anchors were different from the other subscales. Self-control ($\underline{r} = -.26, \underline{p} < .05$), Communality ($\underline{r} = .30, \underline{p} < .05$), and Flexibility ($\underline{r} = -.45, \underline{p} < .05$), significantly correlates with the altered scale. Customer relations, miles per gallon, and general management of drivers, each

significantly correlated with six scales on the CPI. The Flexibility scale on the CPI significantly correlated with five parts of the appraisal form, the total scale, and the altered subscale. However, since a large majority of the correlations were nonsignificant, hypothesis three must be accepted. Also, one must keep in mind the possibility of experiment wise error rate when interpreting the results. This could explain the significant correlations that did occur.

The fourth hypothesis states that there is an absence of relationship between the driver managers' biographical information and behaviorally anchored supervisory ratings. According to Table 3, "number of drivers supervised" correlated with "personal work habits and stress management" ($r = .26, p < .05$) "timeliness" ($r = .34, p < .05$), "Data Accuracy and Timeliness" ($r = .28, p < .05$), "All except Data Accuracy and Timeliness" ($r = .32, p < .05$), and with the total scale ($r = .34, p < .05$). The "number of trucks supervised" scale correlated with "DOT/Safety Compliance" ($r = -.32, p < .05$) and with "Turnover" ($r = -.40, p < .05$).

"Truck driving experience" negatively correlated with "Driver Relations" ($r = -.29, p < .05$), "Utilization" ($r = -.34, p < .05$), "Turnover" ($r = -.25, p < .05$), and "General Management of Drivers" ($r = -.40, p < .05$). A

large majority of scales were nonsignificant, therefore, one must accept hypothesis four.

Table 1

Correlation of the Wesman PCT and the Watson-Glaser CTA
with Ratings for the Total Carriers

	Wesman ($\underline{n} = 48$)	Watson-Glaser ($\underline{n} = 46$)
Driver Relations	.13	.07
Customer Relations	.04	.02
Sales/Marketing Relations	.01	.02
DOT/Safety Compliance	.05	.18
Personal Work Habits and Stress Management	.17	.24*
Miles Per Gallon	.02	.16
Utilization	-.03	.02
Out-of-Route	-.07	.07
Turnover	.09	.11
Availability	.30*	.28*
General Management of Drivers	.04	.17
Timeliness	.09	.15
Data Accuracy	.16	.13
Data Accuracy and Timeliness	.14	.16
All Except Data Accuracy and Timeliness	.15	.22
Total	.16	.22

Note. Wesman scores are verbal scores only.

* $p < .05$

Table 2

Correlation of the CPI with Ratings for All Carriers

	Do	Cs	Sy	Sp	Sa	In	Em	Re	So	Sc
Driver Relations	.13	-.10	.10	-.00	.12	.05	-.01	-.06	.00	-.18
Customer Relations	.12	-.16	.13	.29*	.26*	-.03	.02	-.35*	-.11	-.30
Sales/Marketing Relations	.15	.04	.14	.19	.13	-.02	.22	.10	.02	.07
DOT/Safety Compliance	.18	.14	-.11	-.14	-.07	.12	.06	.12	.22	.20
Personal Work Habits and Stress Management	.18	-.14	.01	-.06	.15	.05	-.03	-.03	-.08	-.20
Miles Per Gallon	.40*	.37*	.35*	.33*	.27	.30*	.25	.09	.08	-.04
Utilization	-.01	.05	.23	.20	.13	.02	.11	-.13	-.06	-.31*
Out-of-Route	.06	-.00	.01	.11	-.00	.12	.20	-.13	-.01	-.14
Turnover	.12	.03	.15	.03	.05	.03	.10	-.06	.05	-.11
Availability	.34*	.01	.26*	.27	.32	.24	.16	-.02	-.04	-.09
General Management of Drivers	.11	-.03	.11	.13	.14	.23	.01	-.29*	-.26*	-.33
Timeliness	.28*	.00	.21	.08	.24	.21	.15	-.04	.13	-.08
Data Accuracy	.16	.14	.19	-.16	.22	.09	.01	-.24*	.05	-.17
Data Accuracy and Timeliness	.25*	-.08	.22	.13	.26*	.17	.10	-.16	.03	-.14
All Except Data Accuracy and Timeliness	.24	-.00	.20	.18	.21	.16	.19	-.16	-.05	-.26*
Total	.25*	-.02	.23	.19	.24	.18	.19	-.17	-.03	-.25*

	Gj	Cm	Wb	To	Ac	Ai	Ie	Py	Fx	F/M
Driver Relations	-.12	.24	.02	-.23	.15	-.19	-.13	-.02	-.30*	-.08
Customer Relations	-.12	.04	-.00	-.29*	-.18	-.20	-.12	-.27*	-.30	.04
Sales/Marketing Relations	.01	.16	.03	.23	.02	.22	-.01	.01	.08	.35*
DOT/Safety Compliance	.24*	.08	.22	.13	.17	.01	-.04	.04	-.19	-.13
Personal Work Habits and Stress Management	-.15	.14	-.06	-.15	.05	-.17	-.26*	-.08	-.21	.11
Miles Per Gallon	.11	.15	.23	-.08	.24	-.18	-.04	-.14	-.35*	-.24
Utilization	-.27*	.23	-.11	-.07	.09	-.06	-.11	-.10	-.14	-.04
Out-of-Route	-.04	.05	.18	-.09	.18	-.16	.00	-.17	-.22	-.28*
Turnover	-.04	.24*	-.01	.01	.24*	-.08	-.13	-.20	-.46*	-.03
Availability	.09	.18	.11	-.03	-.02	.11	.30*	.13	-.18	-.19
General Management of Drivers	-.21	.14	-.10	-.30*	.07	-.33*	-.15	-.16	-.25*	-.08
Timeliness	.03	.22	.11	.07	.22	-.02	-.14	-.03	-.23	.13
Data Accuracy	-.10	.07	.00	-.13	.03	-.13	-.04	-.14	-.19	.20
Data Accuracy and Timeliness	-.04	.16	.07	-.03	.14	-.08	-.11	-.10	-.23	.18
All Except Data Accuracy and Timeliness	-.10	.30*	.05	-.13	.14	-.15	-.11	-.17	-.39*	-.09
Total	-.08	.29*	.06	-.11	.15	-.14	-.11	-.17	-.37*	-.03

Note. See Appendix E for CPI scale definitions.
* $p < .05$

Table 3

Correlation of Biodata with Behavioral Ratings for Total Carriers

	Educational Level	Tenure (Co.)	Tenure Trucking	# of drivers supervised	# of trucks supervised	Truck Driving Experience
Driver Relations	-.10	-.03	-.03	.07	-.02	-.29*
Customer Relations	-.05	-.03	-.17	.10	-.06	-.08
Sales/Marketing Relations	.12	.10	-.02	.16	.21	.09
DOT/Safety Compliance	-.08	-.01	-.005	.21	-.32*	.08
Personal Work Habits and Stress Management	.24	-.20	-.22	.26*	-.001	-.01
Miles Per Gallon	.17	-.23	-.03	.27	.01	-.02
Utilization	-.12	.07	.004	.10	-.15	-.34*
Out-of-Route	-.16	-.17	-.16	.42	-.21	-.07
Turnover	-.20	.06	.09	.21	-.40*	-.25*
Availability	.29*	-.17	-.04	-.09	-.08	-.06
General Management of Drivers	.07	-.12	.09	.16	-.06	-.40*
Timeliness	.18	-.11	-.16	.34*	.02	.02
Data Accuracy	.10	-.01	-.09	.16	.16	.15
Data Accuracy and Timeliness	.16	-.07	-.14	.28*	.10	.10
All Except Data Accuracy and Timeliness	.02	-.10	-.08	.32*	-.20	-.23
Total	.02	-.10	-.10	.34*	-.13	-.15

Note. DOT = Department of Transportation. * $p < .05$

A supplementary factor analysis was performed to investigate the factor structure of the criterion measure. The results appear in Table 4. Using the Varimax Rotated Maximum Likelihood method, four factors were found in the analysis. The first factor includes the Driver Relations, Personal Work Habits and Stress Management, Utilization, and General Management of Drivers subscales from the criterion. This factor was labeled "Management Style," because these scales represent personal characteristics that affect the management style of a driver manager. The second factor includes the Customer Relations, Availability, and Data Accuracy subscales. This factor was labeled "Dependability," because all of these scales represent reliable and dependable behavior. The third factor includes the Sales and Marketing Relations subscale and the fourth factor includes the DOT/Safety Compliance subscale. The last two scales had one subscale that loaded on that factor, so they were named for that subscale. It was decided that four factors would be more efficient than five factors because the third, fourth, and fifth factors had just one subscale that loaded on it. The Chi Square significance levels were the same, .99 for both four and five factors and .94 for three factors. The items that loaded highly on all four factors were omitted.

Table 4

Varimax Rotated Maximum Likelihood Factor Loadings for the
Criterion Measure

Variables	Factors			
	I Mgt. Style	II Depend.	III Sales/ Mktg.	IV Compliance
Driver Relations	.75*	.19	.05	-.06
Customer Relations	.26	.72*	.24	.00
Sales/Marketing Relations	.11	.18	.98*	-.04
Personal Work Habits	.66*	.15	.35	.11
Utilization	.69*	-.02	-.08	-.04
Availability	-.01	.45*	-.08	-.04
DOT/Safety Compliance	-.13	-.04	.01	.78*
Out-of-Route	.42	.18	-.02	.52
Turnover	.50	.04	.01	.39
General Management of Drivers	.67*	.29	-.01	.16
Data Accuracy	.15	.74*	.22	.13
Timeliness	.39	.54	.28	.33
Total Variance accounted for	35.4%	13.6%	12.3%	9.0%
Eigenvalue	4.24	1.63	1.47	1.08
Common Variance accounted for	26.2%	14.3%	8.9%	8.0%

Note. * = inclusion in scale; Depend. = Dependability;
Mktg. = Marketing; Mgt. = Management.

Table 5

Factor Correlations with the Predictors

Variables	Factors			
	I Mgt. Style	II Depend.	III Sales/ Mktg.	IV Compliance
Wesman PCT	.12	.22	.05	.01
Watson-Glaser CTA	.17	.18	.18	.02
Dominance	.12	.25*	.18	.15
Capacity for Status	-.05	-.11	.14	.04
Sociability	.16	.26*	-.11	.14
Social Presence	.10	.32*	-.14	.19
Self-acceptance	.17	.34*	-.07	.13
Independence	.13	.14	.12	-.02
Empathy	.07	.11	.06	.22
Responsibility	-.17	-.27*	.12	.10
Socialization	-.13	-.05	.22	.02
Self-Control	-.33*	-.24	.20	.07
Good Impression	-.22	-.05	.24	.01
Communality	.26*	.13	.08	.16
Well-being	-.07	.05	.22	.03
Tolerance	-.22	-.18	.13	.23
Achievement via Conformance	.09	-.08	.17	.02
Achievement via Independence	-.22	-.09	.01	.22

Table 5--Continued

Variables	Factors			
	I Mgt. Style	II Depend.	III Sales/ Mktg.	IV Compliance
Intellectual Efficiency	-.18	.07	-.04	-.01
Psychological- mindedness	-.11	-.12	-.04	.01
Flexibility	-.25*	-.26*	-.19	.08
Femininity/ Masculinity	-.07	.01	-.13	.35*
Educational Level	.03	.14	-.08	.12
Tenure (Company)	-.10	-.09	-.01	.10
No. of Drivers Supervised	.22	.09	.21	.16
No. of Trucks Responsible for	-.04	.04	-.32*	.21
Truck Driving Experience	-.29*	.02	.08	.09

Note. Mgt. = Management; Mktg. = Marketing; Depend = Dependability.

*p < .05

Discussion

The major issue that the present study attempted to address was the proper selection of future driver managers. It sought to answer the question of whether the test

battery presently being used will be of use to this company in the future, and to select driver managers who will be successful.

The first hypothesis, which states that there would be no relationship between a test of general intelligence and behaviorally based supervisory ratings, was supported, with the exception of one subscale of the supervisory ratings. This leads one to assume that intelligence, beyond some minimum level, is not a requirement for success in the driver management position.

The second hypothesis, which states that there would be no relationship between a measure of critical thinking and behaviorally based supervisory ratings, was supported with similar results, which would lead one to presume that critical thinking is not predictive of driver management success. The Watson-Glaser results were close enough to significance to suggest continuing the use of the test on an experimental basis.

The hypothesis stating that there would be no relationship between a personality measure with behaviorally based supervisory ratings was supported, but it also showed some significant results. The Dominance, Self-control, Communalilty, and Flexibility scale from the personality measure did present interesting results. The more successful driver managers revealed more dominant and

communal characteristics than the less successful driver managers. Perhaps confidence, assertive behavior, and seeing oneself as fitting in with the crowd, facilitate successful performance in a driver manager's job. Also, the more successful driver managers revealed less self-control and less flexibility than the less successful driver managers. This would suggest that someone who expresses his or her emotions and is not flexible would be more successful in the driver management position.

It is suggested that the CPI should be retained as part of the selection process. It correlated well enough with the total ratings to suggest that it would be useful in the future. A simple, unweighted linear composite comprised of the four scales ($Do + Cm - Sc - Fx$) would appear to be promising, but cross validation would be needed.

The hypothesis stating that there would be no relationship between biographical information and supervisory ratings did reveal some significant items. The biodata scales, "number of drivers supervised," and "truck driving experience," seemed to relate to more of the scales than the other items. It could be assumed that the driver managers who were better able to tolerate more stressful situations and had better work habits were assigned more drivers to supervise. The "number of drivers supervised"

subscale significantly correlated with most of the scales and the total scale. Unfortunately, scores on this item are not available for driver managers who have been recently hired or recently promoted.

The driver manager with more trucks to oversee could manifest more safety violations and turnover because they could not supervise all trucks as closely as other driver managers with less trucks to oversee. This information should be incorporated into the performance appraisal system so that driver managers with more trucks to oversee are not penalized. An investigation as to the optimal number of trucks to supervise may be of value to the company in reducing the costs of turnover and safety violations.

The driver managers who had prior truck driving experience have better driver relations are better at organizing their drivers, have less turnover among their drivers, and are better managers of their drivers in general. Fortunately, the "truck driving experience" item is a biographical item that would be available from new hires. This would make further predictive research more convenient.

Because these types of tests have been so successful in the past with the prediction of management success, the lack of positive findings in the present study is puzzling. There are many problems that could have affected the

results. One possible explanation could have been due to the size of the sample (range restriction). Perhaps with a larger sample size the results would have turned out more favorably.

Possibly better and more extensive rater training would have affected the results in a positive way. The raters did seem to score everyone toward the middle of the scale (possibly indicating a central tendency error). Range restriction would reduce the possibility of significant results. This is considered to be a central tendency rater error. Also, it is possible that the scores on the criterion represent a normal distribution.

A factor analysis was conducted for the criterion variables. The analysis uncovered four different factors from the rating scale that supposedly measured four distinct aspects of performance including Management Style, Dependability, Sales and Marketing Relations, and Compliance. These four factors were then reanalyzed for significant relationships. Unfortunately, these new scales only correlated significantly with a few of the predictors. The Dependability factor correlated with more of the subscales than the other three. The results of the factor analysis is tentative considering the small sample size.

These results lead to what recommendations can be suggested to the company involved. First of all, the

experimenter observed that the job analysis techniques were nonstandard. The job analysis was extensive, but it did not have enough structure. Comments about the job and what activities the employees were involved in were written down from observations. Some of the observations included the amount of time taken to complete a task and others did not. Perhaps the job analysis could have been more precise in capturing the important aspects of a driver manager's position. New, standardized methods of analyzing the jobs at the company would greatly facilitate any future validation research and keep the supervisors and employees more aware of the important aspects of their jobs.

Next, the job descriptions for the driver manager position revealed that employees at each of the four sites felt that different activities were important. This confusion about the position could have affected the results of the research. The experimenter suggests a standardized job description throughout the company for each job position.

The testing should be systematic and standardized. This step would facilitate future validation research and create a feeling that all employees are being treated equally.

Another problem within the company is the lack of standardized performance appraisal methods for each position throughout the company and across all carriers. When

employees know what is expected of them in behavioral terms, they are more likely to succeed in the job and supervisors will know what important aspects of the job to stress when training employees. The experimenter suggests developing behaviorally based performance appraisal systems for each position in the company and across all carriers. This again, would facilitate future research projects.

Furthermore, some method of training the people who will be rating employees on the appraisal forms is needed. Training will produce better results for future research and will create a feeling that an attempt is being made to control favoritism and bias.

As a final recommendation to the company, the experimenter suggests choosing other tests for the driver manager position. If time and money permit, an assessment center that simulates a typical driver manager's day could possibly aid in selecting appropriate candidates. Assessment centers can even be used with present employees to correct problems that already exist.

In conclusion, the tests presently being used have proven themselves in the past to predict managerial behavior in other jobs so the problem could either exist in the criterion or in the methods used. More research is needed with the tests and more standardized methods would be useful.

APPENDIX A

DRIVER MANAGER PERFORMANCE RATING FORM

Vice President Operations: _____

Company: _____

Fleet Manager: _____

Driver Manager: _____

Date: _____

Driver Manager Performance Rating Form

Please rate the driver managers using the following scale:

1. Fails to meet performance expectations
2. Meets performance expectations at times
3. Meets performance expectations
4. Exceeds performance expectations
5. Far exceeds performance expectations

(circle one)

- 1 2 3 4 5 Driver Relations--Develops rapport with the drivers in such a way that the driver performs effectively. Listens and talks to drivers to facilitate better understanding and performance and to enhance driver quality of work life.
- 1 2 3 4 5 Customer Relations--Communicates service problems and potential service problems in a timely manner. Understands and is concerned with quality of service.
- 1 2 3 4 5 Sales and Marketing Relations--Effectively represents his/her driver to sales and marketing personnel. Works cooperatively to meet the needs of both drivers and customers.
- 1 2 3 4 5 D.O.T./Safety Compliance--Monitors hours of service and falsifications. Ensures compliance with valid medical and drivers license certificates. Promotes an accident-free working environment. Stresses behavior that is consistent with the D.O.T./safety compliance of the company.

- 1 2 3 4 5 Personal Work Habits and Stress Management--
Prepares worksheets or other aids to use during the day. Checks with night duty for messages upon arrival to work. Follows through on issues left over from the previous day. Plans for the next day. Is seldom absent or tardy. Maintains composure, good judgment, and adequate performance when working under pressure.
- 1 2 3 4 5 Cost Management--Maintains standards set for the fleet of trucks this Driver Manager is responsible for.
- 1 2 3 4 5 MPG
- 1 2 3 4 5 Utilization
- 1 2 3 4 5 Out-of-Route
- 1 2 3 4 5 Turnover
- 1 2 3 4 5 Availability
- 1 2 3 4 5 General Management of Drivers--Sets and enforces reasonable performance standards for drivers. Requires and tracks check calls. Monitors all information pertinent to present or upcoming dispatches. Ensures timely servicing of trucks. Willing to make decisions, anticipate outcomes, and follow through.

When rating with the following two dimensions, keep in mind that a rating of 5 represents standard performance (that is, the performance we hope for from all Driver Managers), 3 represents average performance and 1 represents below average performance.

- 1 2 3 4 5 Data Accuracy--Accurately and completely records information such as location, time, trailer number, condition, and hours of service (logs). Ensures that all driver and load activity is reflected on the computer screens. Monitors fleet screen for accuracy.
- 1 2 3 4 5 Timeliness--Performs and handles daily tasks in a timely fashion, i.e., accepting preassignments, answering/sending messages, service failure reporting, updating ETA's/availability, dispatching trucks and loaded check call.

APPENDIX B

DRIVER MANAGER RATING FORM ANCHORS

Driver Relations

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
<p>Communication with driver is extremely inefficient</p> <p>Presses drivers into impossible commitments</p> <p>Disregards for their needs</p> <p>Doesn't listen or talk to drivers in a productive way</p> <p>Has no rapport with drivers</p>	<p>Communicates with the driver in a somewhat efficient manner</p> <p>Usually compliments drivers when they do a good job</p> <p>Usually listens and talks to drivers in a productive way</p> <p>Has average rapport with the drivers</p>	<p>Communicates with the driver in an efficient manner</p> <p>Consistently listens and talks to drivers in such a way that he/she is able to meet driver needs and achieve good performance</p> <p>Develops an extremely productive rapport with drivers</p> <p>Frequently compliments drivers when they do a good job</p>

Customer Relations

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
Consistently fails to execute the customer commitments	Usually executes the customer commitments	Understands and exceeds the customer expectations of service and executes those commitments with the driver
Rarely communicates service problems or potential service problems	Regularly communicates service problems or potential service problems	Communicates service problems and potential service problems in advance and takes corrective action
Doesn't understand the needs of the customers	Contacts customers when appropriate	Keeps in close contact with customers as appropriate
Never verifies deliveries Does not deal with customers effectively	Generally understands the needs of the customers Regularly verifies deliveries	Follows through on special requirements
Is not courteous when dealing with customers	Usually deals with the customer in a courteous and businesslike fashion	Always deals with the customer in a courteous and businesslike fashion
		Promotes good will with customers

Sales and Marketing Relations

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
Doesn't work as a team with Sales and Marketing	Follows through with Sales and Marketing requests	Regularly communicates with Sales and Marketing to ensure customer expectations and other conditions pertinent to a shipment
Neglects to communicate with Sales and Marketing	Occasionally communicates with Sales and Marketing	Reports available trucks to Sales and Marketing on a proactive basis
Doesn't report available trucks to Sales and Marketing	Usually reports available trucks to Sales and Marketing	Consistently works together to get preassignments before drivers call in
Sales and Marketing has no confidence in Driver Manager such that they do not preassign	Sales and Marketing has average confidence in Driver Manager such that they occasionally preassign	Sales and Marketing have confidence in Driver Manager such that they consistently preassign trucks

D.O.T./Safety Compliance

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
<p>Never follows up on false logs</p> <p>Never plans the drivers next load based on available hours</p> <p>Never informs drivers of upcoming license/physical expirations and/or does not route them to the home terminal before expiration</p> <p>Does not follow up to driver concerning D.O.T./safety after a safety violation has occurred</p> <p>Demonstrates no working knowledge of safety compliance procedures</p>	<p>Follows up on false logs</p> <p>Sometimes plans the drivers next load based on available hours</p> <p>Usually informs driver of upcoming expired license/physical and usually routes them to home terminal before expiration</p> <p>After notification of a D.O.T./safety violation from safety, he/she communicates to the driver the benefits of compliance</p> <p>Demonstrates adequate working knowledge of compliance procedures</p>	<p>Communicates the importance of maintaining an accurate log prior to violations</p> <p>Effectively plans the drivers loads based on available hours</p> <p>Always informs drivers of upcoming expired license/physical and routes them to home terminal before expiration</p> <p>Communicates to the driver the benefits of D.O.T./safety compliance before violations occur</p> <p>Demonstrates superb working knowledge of safety compliance procedures</p>

Personal Work Habits and Stress Management

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
Fighting fires rather than preparing and planning	Usually checks messages	Always checks for messages
Doesn't check messages	Some planning and preparation, mostly short-term plans	Familiarize themselves with issues left over from the previous day
Uses short-term memory rather than computer	Generally understands reports	Prepares to give complete information
Never looks ahead for likelihood of unexpected events	Generally looks ahead for likelihood of unexpected events (geography, weather)	Does today for tomorrow
Never advises night duty of things not accomplished	Usually advises night duty of things not accomplished	Looks ahead for likelihood of unexpected events
Has a pattern of tardiness and absenteeism	Almost never misses a day of work and is seldom tardy	Sets short term goals and develops strategies for meeting goals
		Always advises night duty of things not accomplished
		Is completely dependable, i.e., never misses a day of work and is never tardy

Cost Management

1 3 5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
Consistently fails to meet standards set by company	Consistently meets standards set by company	Consistently exceeds standards set by company
Neglects to analyze driver data	Usually analyzes driver data	Always analyzes driver data

General Management of Drivers

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
<p>Does not set or enforce performance standards</p> <p>Doesn't understand what the reports are for</p> <p>Doesn't give drivers complete information</p> <p>Doesn't monitor information pertinent to dispatches</p> <p>Never solicits information from drivers</p> <p>Get neither drivers needs met nor good productivity</p>	<p>Regularly sets realistic ETA's</p> <p>Usually follows up on driver location</p> <p>Usually monitors information pertinent to dispatch</p> <p>Usually solicits information from drivers</p> <p>Is often forced to choose either drivers needs or good performance</p>	<p>Consistently sets realistic ETA's and communicates to load planner</p> <p>Gives specific instructions</p> <p>Always follows up on driver location</p> <p>Makes sure drivers have check called in and manages exceptions</p> <p>Always monitors all information pertinent to dispatch</p> <p>Continually works ahead of the trucks</p> <p>Sets specific performance expectations when delivery windows are used</p> <p>Constantly solicits information from drivers</p>

Data Accuracy

1

3

5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
10% or more deletions on a regular basis	About 5% deletions on a regular basis	No deletions
Assumes the accuracy of the fleet screen	Occasionally neglects to put correct data in computer	No ETA changes due to data errors
Doesn't accurately and completely record information	Occasionally monitors fleet screen for accuracy	Double checks information
Seldom notices incorrect or missing data and corrects it	Sometimes notices incorrect or missing data and corrects it	Repeats information back to driver for confirmation
		Consistently monitors fleet screen for accuracy
		Consistently notices incorrect or missing data and corrects it

Timeliness

1 3 5

Fails to Meet Performance Expectations	Meets Performance Expectations	Far Exceeds Performance Expectations
<p>Doesn't do computer work while getting dispatch</p> <p>Puts work aside, sometimes forgetting to put it in the computer</p> <p>Rarely follows up on important communication problems immediately</p>	<p>Occasionally records information when it is reported</p> <p>Sometimes doesn't do computer work while receiving dispatch</p> <p>Communicates information as soon as possible</p>	<p>Does computer work when it is reported</p> <p>Communicates information immediately</p>

APPENDIX C

BACKGROUND DATA SHEET

Driver Manager Background Data Sheet

Name: _____ Company: _____

1. Your Month and Year of Birth: Month _____
Year _____
2. Your Sex: (circle one) 1. Male
2. Female
3. Indicate your (circle one) highest level of education:
 1. Did not finish high school
 2. High school graduate
 3. Some college/technical training
 4. College graduate
Major? _____
 5. Graduate Degree
Major? _____
4. How long have you worked: At _____ : _____ Years
_____ Months
 In trucking: _____ Years
_____ Months
5. How many truck drivers do you currently supervise? _____ drivers
6. How many units (trucks) are you currently responsible for? _____ units
7. Have you ever driven (circle one) a truck (over the road) for a living?
 1. Yes _____ Years
_____ Months
 2. No
8. Have you ever served (circle one) in the military?
 1. Yes
 2. No
 Rank: _____
9. Have you ever worked (circle one) in a "service" type job?
 1. Yes
 2. No
10. Marital Status: (circle one)
 1. Single (never married)
 2. Married
 3. Divorced/Separated
 4. Widowed

APPENDIX D
WATSON-GLASER CRITICAL THINKING APPRAISAL
DIMENSION DEFINITION

Watson-Glaser Critical Thinking Appraisal

Scale Definitions

1. Inference: Discriminating among degrees of truth or falsity of inferences drawn from given data.
2. Recognition of Assumptions: Recognizing unstated assumptions or presuppositions in given statements or assertions.
3. Deduction: Determining whether certain conclusions necessarily follow from information in given statements of premises.
4. Interpretation: Weighing evidence and deciding if generalizations or conclusions based on given data are warranted.
5. Evaluation of Arguments: Distinguishing between arguments that are strong and relevant to a particular question at issue.

Source: Watson & Glaser, 1980.

APPENDIX E
CALIFORNIA PSYCHOLOGICAL INVENTORY
DIMENSION DEFINITIONS

CPI Scale Definitions

Scale Name and Implications of Higher and Lower Scores

Do (Dominance)

Higher: confident, assertive, dominant, task oriented

Lower: unassuming, not forceful

Cs (Capacity for Status)

Higher: ambitious, wants to be a success, independent

Lower: unsure of self, dislikes direct competition

Sy (Sociability)

Higher: sociable, likes to be with people, friendly

Lower: shy, feels uneasy in social situations, prefers to keep in the background

Sp (Social Presence)

Higher: self-assured, spontaneous, a good talker, not easily embarrassed

Lower: cautious, hesitant to assert own views or opinions, not sarcastic or sharp-tongued

Sa (Self-acceptance)

Higher: has good opinion of self, sees self as talented, and as personally attractive

Lower: self-doubting, readily assumes blame when things go wrong, often thinks others are better

In (Independence)

Higher: self-sufficient, resourceful, detached

Lower: lacks self-confidence, seeks support from others

Em (Empathy)

Higher: comfortable with self and well-accepted by others, understands the feelings of others

Lower: ill at ease in many situations; unempathic

Re (Responsibility)

Higher: responsible, reasonable, takes duties seriously

Lower: not overly concerned about duties and obligations; may be careless or lazy

So (Socialization)

Higher: comfortably accepts ordinary rules and regulations; finds it easy to conform

Lower: resists rules and regulations; finds it hard to conform; not conventional

Sc (Self-control)

Higher: tries to control emotions and temper; takes pride in being self-disciplined

Lower: has strong feelings and emotions and makes little attempt to hide them; speaks out when angry or annoyed

Gi (Good Impression)

Higher: wants to make a good impression; tries to do what will please others

Lower: insists on being himself or herself even if this causes friction or problems

Cm (Communality)

Higher: fits in easily; sees self as a quite average person

Lower: sees self as different from others; does not have the same ideas, preferences, etc., as others

Wb (Well-being)

Higher: feels in good physical and emotional health; optimistic about the future

Lower: concerned about health and personal problems; worried about the future

To (Tolerance)

Higher: is tolerant of others' beliefs and values, even when different from or counter to own beliefs

Lower: not tolerant of others; skeptical about what they say

Ac (Achievement via Conformance)

Higher: has strong drive to do well; likes to work in settings where tasks and expectations are clearly defined

Lower: has difficulty in doing best work in situations with strict rules and expectations

Ai (Achievement via Independence)

Higher: has strong drive to do well; likes to work in settings that encourage freedom and individual initiative

Lower: has difficulty in doing best work in situations that are vague, poorly defined, and lacking in clear-cut methods and standards

Ie (Intellectual Efficiency)

Higher: efficient in use of intellectual abilities; can keep on at a task where others might get bored or discouraged

Lower: has a hard time getting started on things and seeing them through to completion

Py (Psychological-mindedness)

Higher: more interested in why people do what they do than in what they do; good judge of how people feel and what they think about things

Lower: more interested in the practical and concrete than the abstract; looks more at what people do than what they feel or think

Fx (Flexibility)

Higher: flexible; likes change and variety; easily bored by routine life and everyday experience; may be impatient and even erratic

Lower: not changeable; likes a steady pace and well-organized life; may be stubborn and even rigid

F/M (Femininity/Masculinity)

Higher: sympathetic, helpful; sensitive to criticism; tends to interpret events from a personal point of view; often feels vulnerable

Lower: decisive, action-oriented; takes the initiative; not easily subdued; rather unsentimental

APPENDIX F
RATER TRAINING FOR DRIVER MANAGER
PERFORMANCE RATING FORM

Rater Training for Driver Manager
Performance Rating Form

READ THIS FIRST

Job incumbents display relative strong points and weak points in their performance, and your ratings should reflect those strengths and weaknesses. Some common rater errors that should be avoided are listed below.

1. Rate behavior not ability--rate what the driver manager does not what you think he/she is capable of doing.
2. Just because a person is strong in one area doesn't mean they are strong in all areas. This is a classification error called "Halo" error. Don't let one attribute affect how you rate all other areas.
3. People are naturally different--your ratings should be, too. The impact of lumping everyone around one part of the scale will not let your star performers stand out, and will give your poor performers a place to hide--use the whole scale.

*Please remember that these ratings are for research purposes only.

APPENDIX G
COVER LETTER TO DRIVER MANAGERS

Memorandum
Fort Worth, Texas
May 4, 1989

TO: Driver Managers/Driver Coordinators

FROM:

RE: Driver Manager Research Project

is currently researching the Driver Manager position in the hopes that when hiring future Driver Managers, we will be able to predict if a job applicant will be successful. To get at this goal we are seeking answers to the following questions:

1. Who are our current driver managers and what kinds of life experiences are typical among the driver managers?
2. Are there similarities in the personalities of the driver managers?

To complete this project we need your help. First, there is a background data sheet enclosed with this package. It asks a number of questions about you such as your educational background, whether or not you have ever driven over the road, etc.

Next, if you have not had the opportunity to take part in the Management Assessment process previously, this will be your chance. The enclosed inventories are widely used in this type of research and will NOT be put in your personnel file.

You will never be identified individually in this study. You will, however, receive a personal copy of the study summary of the results if you take part.

Because there are not many driver managers, we really do need your help. However, if you have strong objections to taking part in this project, you can decline to take part.

If you have any questions or concerns, please feel free to talk to your local Vice President-Operations, or call me in Fort Worth at

Thank you for your help.

Director-Human Resources,

APPENDIX H
MEANS AND STANDARD DEVIATIONS OF
DRIVER MANAGER ON THE
WESMAN PCT AND THE WATSON-GLASER CTA

Means and Standard Deviations on Two Ability Tests

	Groups				
	A (<u>n</u> =15)	B (<u>n</u> =3)	C (<u>n</u> =18)	D (<u>n</u> =12)	Total (<u>n</u> =48)
	The Wesman PCT				
Mean	22.47	16.33	17.06	20.25	19.50
SD	6.66	4.93	5.92	7.86	6.88
	Watson-Glaser CTA				
Mean	58.43	51.00	49.00	56.25	53.89
SD	10.12	5.00	9.58	9.76	10.20

Note. A, B, C, and D are carriers that represent different areas of the country.

APPENDIX I
MEANS AND STANDARD DEVIATIONS OF
DRIVER MANAGER ON THE
CALIFORNIA PSYCHOLOGICAL INVENTORY

California Psychological Inventory Mean and Standard
Deviation in Standard Scores

	Males (n = 39)		Females (n = 9)	
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
Do	60.59	11.68	57.44	10.71
Cs	51.82	8.52	46.67	10.02
Sy	53.54	9.67	53.44	7.63
Sp	48.19	11.47	50.00	7.98
Sa	52.73	9.73	56.67	9.51
In	52.97	7.35	53.78	7.90
Em	51.65	10.91	49.44	7.47
Re	51.22	8.98	46.78	7.36
So	51.03	10.20	48.56	7.65
Sc	51.08	9.44	47.78	6.94
Gi	53.24	11.18	50.22	9.09
Cm	56.19	5.74	55.44	5.25
Wb	51.11	7.88	48.00	5.27
To	47.16	7.28	45.22	8.83
Ac	53.89	9.11	51.11	8.04
Ai	48.65	6.51	46.89	4.48
Ie	47.95	8.62	46.44	5.27
Py	51.89	8.60	45.56	8.02
Fx	43.19	7.26	44.00	7.05
F/M	46.68	10.15	46.89	13.55

REFERENCES

- Abt, L. E. (1949). A test battery for selecting technical magazine editors. Personnel Psychology, 2, 75-91.
- Anastasi, A. (1982). Psychological testing (5th ed.). New York: MacMillan.
- Ash, P. (1960a). Validity information exchange, No. 13-05. Personnel Psychology, 13, 454.
- Ash, P. (1960b). Validity information exchange, No. 13-06. Personnel Psychology, 13, 454.
- Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. Journal of Applied Psychology, 66, 1-6.
- Bass, B. M., & Barrett, G. V. (1984). People, work, and organizations (2nd ed.). Boston: Allyn & Bacon.
- Bemis, S. E. (1968). Occupational validity of the General Aptitude Test Battery. Journal of Applied Psychology, 52, 240-249.
- Bogard, H. M. (1960). Union and management trainees--A comparative study of personality and occupational choice. Journal of Applied Psychology, 44, 56-63.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60, 556-560.

- Cascio, W. F. (1987). Applied psychology in personnel management (3rd ed.). New Jersey: Prentice-Hall, Inc.
- Crites, J. O. (1972). Review of test: Watson-Glaser Critical Thinking Appraisal. In O. K. Burson (Ed.), The sevent mental measurements yearbook (Vol. 2). New Jersey: Gryphon Press.
- Cronbach, L. J. (1959). Review of the California Psychological Inventory. In O. K. Buros (Ed.), The fifth mental measurements yearbook. New Jersey: Gryphon Press.
- Freyd, M. (1923). Measurement in vocational selection: An outline of research procedures. Journal of Personnel Research, 2, 377-385.
- Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.
- Goodstein, L. D., & Schrader, W. J. (1963). An empirically-derived managerial key for the California Psychological Inventory. Journal of Applied Psychology, 47, 42-45.
- Gough, H. G. (1987). Manual for the California Psychological Inventory (rev. ed.). California: Consulting Psychologists.
- Grant, D. L. (1980). Issues in personnel selection. Professional Psychologist, 11, 369-384.

- Guion, R. M. (1965). Personnel testing. New York: McGraw Hill.
- Guion, R. M. (1976). Recruiting, selection, and job placement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 777-828). Chicago: Rand-McNally.
- Guion, R. M. (1987). Changing views for personnel selection research. Personnel Psychology, 73, 199-213.
- Guion, R. M., & Cranny, C. J. (1982). A note on concurrent and predictive validity designs: A critical reanalysis. Journal of Applied Psychology, 67, 239-244.
- Guion, R. M., & Gibson, W. M. (1988). Personnel selection and placement. Annual Review of Psychology, 39, 349-374.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. Personnel Psychology, 18, 135-164.
- Hakstian, R. A., Woolsey, L. K., & Schroeder, M. L. (1987). Validity of a large-scale assessment battery in an industrial setting. Educational and Psychological Measurement, 47, 165-178.
- Helmstadter, G. C. (1972). Review of test: Watson-Glaser Critical Thinking Appraisal. In O. K. Buros (Ed.), The seventh mental measurements yearbook (Vol. 2). New Jersey: Gryphon Press.

- Heneman, H. G. (1975, June). Research roundup. Personnel Administrator, 64.
- Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological intervention on employee performance and work force productivity. American Psychologist, 38, 473-478.
- Kelly, E. L. (1965). Review of the California Psychological Inventory. In O. K. Buros (Ed.), The sixth mental measurements yearbook. New Jersey: Gryphon Press.
- Kleiman, L. S., & Faley, R. H. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. Personnel Psychology, 38, 803-833.
- Labor Letter (1984, August 28). Wall Street Journal, 1.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Latham, G. P., Wexley, K. N. Pursell, E. D. (1985). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- Lee, R., Miller, K. J., & Graham, W. K. (1982). Corrections for restriction of range and attenuation in criterion-related validation studies. Journal of Applied Psychology, 67, 637-639.

- Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971).
Research design and validity assessment. Personnel Psychology, 24, 247-274.
- Lent, R. H., Levin, L. S., Aurbach, H. A. (1971).
Predictors, criteria, and significant results.
Personnel Psychology, 24, 519-533.
- Linn, R. L., Harnisch, D. L., & Dunbar, S. B. (1981).
Corrections for range restriction: An empirical
investigation of conditions resulting in conservative
corrections. Journal of Applied Psychology, 66, 655-
663.
- Magargee, E. I. (1972). The California Psychological
Inventory handbook. London: Jossey-Bass, Inc.
- Modjeski, R. B., & Michael, W. B. (1983). An evaluation
by a panel of psychologists of the reliability and
validity of two tests of critical thinking. Educational
and Psychological Measurement, 43, 1187-1197.
- Moskal, B. S. (1989). Shippers in for a rude awakening.
Industry Week, 38, 20-21.
- Munsterberg, H. (1913). Psychology and Industrial
Efficiency. New York: Riverside Press.
- Parisher, D., Rios, B., & Reilly, R. R. (1979).
Psychologists and psychnological services in urban
police departments: A national survey. Professional
Psychologist, 10, 6-7.

- Rawls, D. J., & Rawls, J. R. (1968). Personality characteristics and personal history data of successful and less successful executives. Psychological Reports, 23, 1032-1034.
- Ryan, A. M., & Sackett, P. R. (1987). A survey of individual assessment practices by I/O psychologists. Personnel Psychologist, 40, 455-488.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing--old theories and new research findings. American Psychologist, 36, 1128-1137.
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.
- Spooner, L. (1985). Personality tests help when hiring. Savings Institution, 106, 158-162.
- Sussman, M., & Robertson, D. U. (1986). The validity of validity: An analysis of validation study designs. Journal of Applied Psychology, 71, 461-468.
- Taylor, G. S., & Zimmerer, T. W. (1988). Personality tests for potential employees: More harm than good. Personnel Journal, 67, 60-64.

- Tziner, A. (1984). A fairer examination of rating scales when used for performance appraisal in a real organization setting. Journal of Occupational Behavior, 5, 103-112.
- Watson, G., & Glaser, E. M. (1980). Manual for the Watson-Glaser Critical Thinking Appraisal. New York: The Psychological Corporation.
- Wesman, A. G. (1965). Manual for the Wesman Personnel Classification Test (Rev. ed.). New York: The Psychological Corporation.
- Westbrook, B. W., & Sellers, J. R. (1967). Critical thinking, intelligence, and vocabulary. Educational Psychological Measurement, 27, 443-446.
- Wysocki, B. (1981, February 25). More companies try to spot leaders early, guide them to the top. Wall Street Journal.
- Zdep, S. M. (1969). Intragroup reinforcement and its effect on leadership behavior. Organizational Behavior and Human Performance, 4, 284-298.