An Integration of Biographical Data and Personality Research Through Sherwood Forest Empiricism: Robbing From Personality to Give to Biodata.

Bruce W. Davis

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/6477
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
AN INTEGRATION OF BIOGRAPHICAL DATA AND PERSONALITY RESEARCH THROUGH
SHERWOOD FOREST EMPIRICISM:
ROBBING FROM PERSONALITY TO GIVE TO BIODATA

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by

Bruce W. Davis
B.S., San Diego State University, 1990
M.A., Louisiana State University, 1992
August 1997
ACKNOWLEDGMENTS

This dissertation was prepared under the direction of Dr. Timothy Buckley. I am extremely grateful for his time, energy, expertise, and assistance. I would also like to thank Dr. Irving M. Lane who provided me with the initial opportunity to attend L.S.U. and since that time has consistently offered excellent guidance, advice, and support. I also appreciate the valuable comments and suggestions provided by committee members Dr. Kevin W. Mossholder and Dr. Joe C. Witt. Special thanks to Dr. Dirk D. Steiner who assisted with earlier drafts of the dissertation proposal and Dr. Craig J. Russell for sparking an early interest in biographical data. I would also like to thank Lee Bebb, Lynn Bride, and Patt Zamary from the County of San Diego for their support, encouragement, and commitment. Finally, I would like to thank my wife, Suzanne for providing inspiration, offering words of encouragement and understanding, and enduring the necessary sacrifices to help me complete this dissertation.
## TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ................................................................................................................... ii

**LIST OF TABLES** ............................................................................................................................... vi

**LIST OF FIGURES** ............................................................................................................................. viii

**ABSTRACT** ......................................................................................................................................... ix

**INTRODUCTION** . ................................................................................................................................ 1

**ADVANCES IN PERSONALITY RESEARCH** ................................................................. 2

Personality and the Big Five ................................................................................................................. 2

Summary ................................................................................................................................................. 6

Criticism of Personality Research..................................................................................................... 7

Person vs. Situation ............................................................................................................................... 7

Personality Measures as Self-Report ................................................................................................. 11

Interactional Approach ......................................................................................................................... 12

Organizations as Situations.................................................................................................................... 13

Test Administration as a Situation .................................................................................................... 15

Summary ................................................................................................................................................. 15

Item Level of Analysis .......................................................................................................................... 16

**ADVANCES IN BIODATA** .......................................................................................................... 18

Biodata Theory ........................................................................................................................................ 19

Developmental-Integrative Model ......................................................................................................... 20

Ecology Model ....................................................................................................................................... 20

Social Identity Theory .......................................................................................................................... 21

Commonalities with Personality Theories ........................................................................................... 22

Socioeconomic Correlates ..................................................................................................................... 23

Proximal Variables and Differential Experience .............................................................................. 23

Incumbent Samples and Item Development .................................................................................... 25

Comparison at Different Levels ............................................................................................................ 26

Theoretical Level .................................................................................................................................... 26

Item Development .................................................................................................................................. 27

Item Characteristics ............................................................................................................................... 27

Social Desirability ................................................................................................................................. 29

Response Similarities ............................................................................................................................ 30

Summary ................................................................................................................................................ 31

Biographical Data Scaling Techniques ................................................................................................ 31

Rational Scaling ..................................................................................................................................... 32

Factorial Scales ....................................................................................................................................... 32

Empirical Scales ..................................................................................................................................... 33

Comparison of Scaling Techniques .................................................................................................... 34

Rainforest Empiricism ............................................................................................................................ 36

Summary ................................................................................................................................................ 37

**INTEGRATION OF THEORIES AND HYPOTHESES** .......................................................... 37

Importance of the Research .................................................................................................................. 38

Biographical Scoring of Personality Measures ................................................................................... 40

Big Five and Absenteeism ...................................................................................................................... 41

Conscientiousness and Absenteeism .................................................................................................. 45
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF TABLES

1. Sample Two Characteristics ................................................................. 54
2. Sample One Demographic Characteristics ........................................... 59
3. Mean Differences for PEPI Scales ......................................................... 59
4. Correlations Between Personality and Biodata Scales with Absenteeism Criterion ............... 62
5. Significance Tests and Confidence Intervals (CI) for Biodata and Personality ..................... 63
6. Regression Results for Absenteeism Variable ......................................... 64
7. Personality and Biodata Scored Dimension Scores for High and Low Absenteeism ................. 66
9. Correlations between Personality and Biodata PEPI Scales with Turnover Criterion ................. 72
10. Personality and Biodata PEPI Scored Dimension Scores for Stayers and Quitters .................. 73
11. HPI Sample Characteristics as a Percentage of Total .................................. 80
12. Group Comparison for Absenteeism or HPI Scale Differences .............................. 80
13. Personality and Biodata Correlations with Absenteeism ..................................... 81
14. HPI Biodata and Personality Comparisons ........................................... 82
15. Regression Results for HPI Biodata and Personality with Absenteeism Variable .................. 83
16. HPI Personality and Biodata Scored Dimension Scores for High and Low Absenteeism ............ 84
18. Reliabilities of Personality and Biodata Scales and Percentage of Items Used for 1991 - 1996 ... 91
19. Sample Demographic Characteristics for Sample Three ...................................... 92
20. CPI Personality Mean Scale and Absenteeism Comparisons for Hold-Out and Development Groups ................................................................. 93
21. CPI Biodata Mean Scale Comparisons for Hold-Out and Development Groups .................. 94
22. Correlations Between Biodata and Personality Scales and Absenteeism ............................ 96
23. Significance Tests and Confidence Intervals (CI) for Biodata and Personality ..................... 99


<table>
<thead>
<tr>
<th>Figure Range</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a - 1d.</td>
<td>Sample One Absenteeism PAMS Profiles for Personality and Biodata</td>
<td>67</td>
</tr>
<tr>
<td>2a - 2d.</td>
<td>Sample One PAMS Profiles for Personality and Biodata</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Dimension Scores for Turnover</td>
<td></td>
</tr>
<tr>
<td>3a - 3d.</td>
<td>Sample Two PAMS Absenteeism Profiles for Personality and Biodata</td>
<td>85</td>
</tr>
<tr>
<td>4a - 4d.</td>
<td>PAMS Profiles for Personality and Biodata for Aggregate Sample</td>
<td>120</td>
</tr>
<tr>
<td>5a - 5d.</td>
<td>PAMS Profiles for Personality and Biodata for the Sheriff's Department</td>
<td>121</td>
</tr>
<tr>
<td>6a - 6d.</td>
<td>PAMS Profiles for Personality and Biodata for the Probation Department</td>
<td>122</td>
</tr>
</tbody>
</table>
ABSTRACT

Archival data was used to assess the effectiveness of incorporating biodata scoring into extant personality measures. Personality and biodata theories were briefly reviewed and several commonalities were noted. Hypotheses were developed for Conscientiousness, Emotional Stability, Agreeableness, and Openness to Experience. The first three dimensions were expected to predict absenteeism and all four were expected to predict turnover, and empirical scoring was expected to increase predictive validity. Data were collected from three samples (school bus drivers, bus drivers, and law enforcement) and three different personality measures were scored using England's (1971) vertical percent method. Results showed that two of the three samples did not produce significant correlations between personality and two criterion measures. Only the correlation between Conscientiousness and absenteeism (r = .13) for the law enforcement sample was similar to Barrick and Mount's (1991) results. Personality profiles for high and low absenteeism employees, and stayers and leavers were compared, but the profiles did not contain the hypothesized elevations. Biodata scoring improved some correlations, but inconsistently, for law enforcement employees. Biodata scoring for this sample was most useful for predicting turnover. Correlations for the other samples were small and insignificant. Whereas biodata scoring provided incremental validity over the personality scales, the results were inconsistent and therefore did not support the hypotheses. Further, profiles developed through biodata scoring diverged from those developed via personality scoring. The interpretation of biodata profiles was difficult due to possible changes in scale meaning due to scoring. The overall component of personality and biodata profiles were compared and it was shown that sometimes the biodata overall component explained incremental variance. Limitations of the study included using incumbents versus applicants, a lack of specificity of the criterion measures, and an inability to compare the results from the three samples. Future studies are needed to determine possible applicant-incumbent differences for personality and biodata measures and determine possible changes in the structure of personality tests caused by empirical scoring. These results provide
disconfirming evidence for personality research and support the need to reconsider the situational specificity hypothesis.
INTRODUCTION

Developing personnel selection instruments that are useful and valid is one of the principle domains of industrial and organizational psychology. Toward this end, researchers have devoted considerable effort and resources to the design of selection instruments and systems that achieve high validity, high utility, and minimize adverse impact (Hunter & Hunter, 1984; Schmitt & Robertson, 1990). To date, reviews of selection literature suggest that the cognitive ability test is the best predictor of job performance (e.g., Hunter & Hunter, 1984; Muchinsky, 1986; Ones, Viswesvaran, & Schmidt, 1993). However, concerns over the potential adverse impact of cognitive ability tests have resulted in continued efforts to develop alternative methods (e.g., Schmitt & Noe, 1986). Recently, alternatives such as personality instruments and biographical data have become leaders in the quest towards this new frontier (Fleishman, 1988). But, to become leaders, both have had to address criticisms. In the case of personality, the criticism has centered around low validity coefficients (e.g., Guion & Gottier, 1965; Reilly & Chao, 1982). For biographical data (or biodata), the criticism has been that the approach is void of theory and is simply "dustbowl empiricism" (e.g., Mumford & Owens, 1987). Researchers in both areas have responded to these criticisms and have made several significant advancements (e.g., Ones et al., 1993). Yet, progress in personality testing and biodata has resulted in less distinct boundaries between the two approaches. This overlap should be construed as a benefit because combining the strengths of each field can advance personnel selection research and application (e.g., Russell, 1994; Tenopyr, 1994).

The purpose of this work is to discuss how personality and biodata can be combined to develop a useful technique for personnel selection. To achieve this goal, this paper is divided into three sections. The first section presents recent findings in personality research and discusses criticisms of paper and pencil personality inventories. The second section presents theoretical progress made with biodata with an emphasis placed on the overlap with personality measures. The different approaches to scaling biodata instruments are also presented in this section. The third section presents an applied...
technique to combine the two approaches to predict absenteeism and turnover. In this section several testable propositions are presented to test the technique. In short, the goals of this project are consistent with the current philosophy of industrial and organizational science as stated by Campion (1996) "...to do what our profession is all about: Use science to solve practical problems, and in so doing learn something more about the science" (p. 9).

ADVANCES IN PERSONALITY RESEARCH

Personality and the Big Five

Historically, personality constructs have achieved limited success in predicting performance. Authors such as Ghiselli and Barthol (1953) and Guion and Gottier (1965) were among the first to report the poor predictive power of personality measures. Schmitt, Gooding, Noe, and Kirsch (1984), in a quantitative review of personality test literature, reported validities ranging from .12 for turnover to .27 for wages, with an average of .15 for all criteria. Based in part on the low validity coefficients, many researchers abandoned the use of personality measures for selection (Hogan, 1991).

Within the past decade, however, there has been a revitalization in personality research. This revitalization has been fueled by the discovery of the Big Five taxonomy of personality (e.g., Costa & McCrae, 1985; Digman, 1990; Digman & Inouye, 1986; Goldberg, 1990). Whereas researchers have not reached consensus about the labeling of the dimensions of personality (Block, 1995), many have adopted the terminology proposed by Digman (1990) and Goldberg (1990) who maintain that Extraversion, Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience provide the best labels for the five factors (e.g., see Hogan, 1991; Revelle, 1995 for recent reviews).

Traits associated with Extraversion, sometimes known as Surgency, include being sociable, gregarious, assertive, talkative, and active (McCrae & Costa, 1985). However, Hogan (1986) splits this dimension into two components: (1) Ambition, which includes initiative, surgency, ambition, and impetuousness and (2) Sociability, defined by traits such as sociable, exhibitionist, and expressive. The second factor, Emotional Stability, is associated with being anxious, depressed, angry, embarrassed.
emotional, worried, and insecure (McCrae & Costa, 1985). Traits associated with Agreeableness include being courteous, flexible, trusting, good-natured, cooperative, forgiving, soft-hearted, and tolerant. Conscientiousness is comprised of a dependability component, associated with being careful, thorough, responsible, organized, and planful (Hogan, 1986); and a volitional component, associated with being hardworking, achievement-oriented, and persevering (Digman, 1990). Finally, traits believed to be associated with the Openness to Experience dimension include being imaginative, cultured, curious, broad-minded, intelligent, and artistically sensitive (Digman, 1990).

Support for the structure of the five factor taxonomy has been demonstrated in a number of studies (Hogan, 1991; Revelle, 1995). These studies have employed a variety of methods, including self-ratings and peer ratings, adjective checklists, and short phrase items, and various methods of analysis (Wiggins & Pincus, 1992). In addition, these studies have been performed with a variety of samples including cross cultural ones (Revelle, 1995). The success of the Big Five taxonomy in other domains of psychology has led industrial and organizational psychologists to refocus their personality-based selection research (Hogan, 1991). Using the Big Five framework, many researchers have begun a search to identify the personality constructs associated with various criteria in a number of occupations (e.g., Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

Findings from recent meta-analytic research suggest that dimensions of the Big Five provide promise for those interested in using personality based measures for selection (e.g., Barrick & Mount, 1991; Robertson & Kinder, 1993; Tett, Jackson, & Rothstein, 1991). However, results from two meta-analyses have reached divergent conclusions regarding the relative importance of the five factors. On the one hand, Tett et al. (1991) found the corrected correlation between Agreeableness, Openness to Experience, Emotional Stability, Conscientiousness, and Extraversion and job proficiency to be .33, .27, .22, .18, and .16, respectively. In contrast, Barrick and Mount found Conscientiousness to be the most important predictor of job performance. They obtained corrected correlations between Conscientiousness and job proficiency, training proficiency, and personnel data that ranged from .20.
to .23 for all occupations. They also found that the corrected relationship between Extraversion and job proficiency, training proficiency, and personnel data to be .20, but only for manager and sales occupations. Finally, whereas the remaining dimensions of personality were significant for some occupations and some criterion measures, the correlations did not approach the magnitude of those obtained with Conscientiousness. The main point is that the two meta-analyses differed in the relative importance and rank order of Agreeableness, Conscientiousness, Extraversion, Openness to Experience, and Emotional Stability.

Ones, Mount, Barrick, and Hunter (1994) attempted to resolve the studies' differences and questioned the findings of Tett and associates on three accounts. First, Ones et al. stated that Tett et al.'s meta-analysis was not comprehensive and excluded many important studies. Second, Ones et al. claimed that the procedure used to assign personality scales from specific inventories to personality constructs differed from the procedure used by Barrick and Mount (1991). Finally, Ones et al. questioned Tett and associates' method of using the absolute value of correlations. Ones and associates maintained that this served to inflate the corrected correlations. In a subsequent rebuttal, Tett, Jackson, Rothstein, and Reddon (1994) addressed the concerns of Ones et al., reanalyzed the data, and reached similar conclusions to those obtained in the original Tett et al. (1991) study. Nonetheless, Tett et al.'s use of absolute values for correlations instead of averaging the signs remained unresolved as Tett et al. (1994) continued to employ the procedure. In spite of these differences in methodology, the original meta-analyses of Barrick and Mount (1991), and Tett et al. (1991), obtained similar correlations between Conscientiousness and job proficiency (.23 and .18, respectively), providing further support for the importance of this dimension (Ones et al., 1994).

Recent studies affirm the contention that Conscientiousness offers the most promise as a predictor of motivational factors of job performance (Barrick & Mount, 1991; Hough et al., 1990; Schmidt & Hunter, 1992). The identification of motivational variables that predict job performance has been a highly sought after goal of practitioners and theorists, but to date has proven somewhat elusive.
(Kanfer, 1991). This is an especially important goal in light of the fact that cognitive ability tests, which have the highest predictive validity for job performance (e.g., Hunter & Hunter, 1984), are not intended to assess motivational constructs (Robertson, 1993). Therefore, continued investigations using Conscientiousness and the other Big Five constructs may hold a key to uncovering motivational factors. Along these lines, Barrick, Mount, and Strauss (1993) and Schmidt and Hunter (1992) stated that Conscientiousness is the most important trait motivation variable in personnel psychology. To provide empirical support for this position, Barrick et al. (1993) demonstrated how Conscientiousness may influence job performance via the goal-setting process. They found that sales representatives who received high scores on the Conscientiousness dimension were more likely to set goals and more likely to be committed to goals. These goals resulted in higher sales volume and higher supervisory ratings of job performance.

In addition to Conscientiousness being an important construct in the prediction of effective job performance, there is research to indicate that it is equally effective in predicting counterproductive performance. Counterproductive behavior includes such manifestations as malingering, equipment damage, insubordination, rule infractions, theft (e.g., Hogan & Hogan, 1989), and absenteeism (Harrison & Shaffer, 1994; Johns, 1994b). Studies by Dunn, Mount, Barrick, and Ones (1995) and Hough et al. (1990) provide empirical support for the relationship between Conscientiousness and counterproductive behavior. In a study of 84 personnel managers, Dunn et al. found that general mental ability and Conscientiousness were the most important attributes related to applicants' hirability and that Emotional Stability, Conscientiousness, and Agreeableness were the most important attributes related to counterproductivity. They used policy capturing to demonstrate that managers were likely to identify levels of Conscientiousness as the best predictor of the employees' counterproductive behavior. Similarly, Hough et al. (1990) reported in their meta-analysis of personality instruments an average correlation between dependability (Conscientiousness in the Big
Five taxonomy) and counterproductive behavior of -.28 and between achievement and counterproductive behavior of -.35.

The relationship between Conscientiousness and counterproductive behavior is a fruitful area of research both from a theoretical standpoint and a practical one. Due to the cost of counterproductivity (Cascio, 1991), it is of paramount importance for organizations to identify and to screen these employees through its selection practices (Hogan & Hogan, 1989; Martin & Terris, 1991). One component of counterproductivity that many organizations would like to reduce is absenteeism (Johns, 1994b; Martocchio, 1992). This is especially true in organizations where being present or on time is critical for business effectiveness (Martocchio, 1992). A second component of counterproductivity that may be an outcome of this type of behavior is employee turnover (Hogan & Hogan, 1989). Although turnover may occur for a number of reasons (e.g., Mobley, Griffeth, Hand, & Meglino, 1979), excessive absenteeism, rule infractions, and insubordination are all likely to result in terminations (e.g., Hollenbeck & Williams, 1986). Researchers have attempted to predict the propensity or intention to engage in deviant behavior, and hence reduce turnover costs (e.g., Cascio, 1991) with personality-based integrity tests (Sackett, Burris, & Callahan, 1989). However, measures of Conscientiousness map many of the same constructs as do integrity tests and also are likely to be valid in the prediction of counterproductive behavior (e.g., Murphy & Lee, 1994).

Summary. There is an accumulation of evidence that suggests that Conscientiousness is effective in predicting motivational aspects of job performance. Although the main thrust of previous research has been in predicting job performance and proficiency ratings (e.g., Barrick & Mount, 1991; Tett et al., 1991), there is a growing research base that suggests that Conscientiousness is equally effective in predicting other organizational outcomes such as absenteeism and turnover (e.g., Barrick & Mount, 1996; Hogan & Hogan, 1989).
Criticism of Personality Research

Even though advances have been made in personality research, there are still some criticisms that must be addressed. One is the lingering debate over the person versus the situation in explaining behavior. A second is the use of paper and pencil measures to explain behavior. Each of these will be discussed in turn.

**Person vs. Situation.** Despite the support for the role of the Big Five in job performance and motivation, there continues to be a debate over the importance of person versus situational variables in explaining the relationship (e.g., Mischel, 1968). As indicated in the studies reviewed below, it is likely to come as no surprise that both variables play a significant role in the prediction of behavior. In terms of the person, research by Hershberger, Plomin, and Pedersen (1995) supports previous findings of a genetic component to personality (e.g., Bouchard & McGue, 1990; Plomin, Chipuer, & Loehlin, 1990; Plomin & Rende, 1991). They found that on average, 15% of the variance in personality scores was attributable to genetic influence, 5% to shared-rearing environmental influence, and 80% to non-shared environmental influence and error. To further bolster the role of the genetic component of personality, Hershberger et al. found that personality is relatively stable over the life course.

Despite the evidence for the stability and heritability of personality, there remains the question posed by Mischel (1968) regarding the consistency of behavior across situations. Weiss and Adler (1984) contended that researchers should not mandate that behavior be consistent across every situation. Instead, they argued that behavior have some degree of coherence in various contexts, as it is unlikely that each person would act the same way in every situation (Revelle, 1995). According to Weiss and Adler, one should be concerned with the relative influence of personality across situations rather than the absolute influence.

Weiss and Adler's (1984) conceptualization may help explain some of the conflicting findings regarding the cross-situational consistency of personality. For example, Funder, Kolar, and Blackman
(1995) found that ratings of the Big Five personality dimensions provided by the self, college acquaintances, hometown acquaintances, parents, and strangers were significantly correlated. Specifically, they found that the average correlation across raters and situations averaged .27 with a range of .24 to .54. Because ratings were obtained in a variety of contexts, the authors were able to conclude that personality was consistent across situations.

In contrast, a study by Mount, Barrick, and Perkins-Strauss (1994) found that personality trait ratings were differentially predicted by supervisors, co-workers, and self ratings. Their study was designed to examine the ability of personality trait ratings made by observers to predict job performance ratings. Mount et al. found that personality ratings made by supervisors and co-workers were the best predictor of performance ratings. For example, supervisor ratings of the Big Five dimensions were all significantly related to performance ratings, and co-worker ratings were related for three of the factors (i.e., Conscientiousness, Extraversion, and Openness to Experience). Conversely, for self-ratings, Conscientiousness was the only correlate of job performance. Mount et al.'s explanation for the finding that co-workers and supervisors made different personality evaluations than the self-ratings made by workers was that co-workers and supervisors assigned ratings using a specific frame-of-reference related to a work context. In Mount et al.'s study, observers' assessments (i.e., co-workers and supervisors) were restricted to job-related observations, or strong situations in which behavioral choices can be constrained, whereas self-ratings made by workers were presumably determined across situations. This differs from Funder and associate's findings because participants and observers were able to rate personality across a variety of contexts thus providing support for Weiss and Adler's (1984) "coherency" argument. In short, Mount et al.'s findings that observers' ratings of personality were better predictors of job performance than were self-assessments suggests that personality instruments tailored to the work situation may be better predictors than those instruments designed to measure personality across situations.
The studies reviewed above suggest that the context or situation in which ratings are made is important and may influence the predictive validity of personality measures (Cattell, 1986; Merydith & Wallbrown, 1991). Yet restricting personality ratings to work-related situations is probably not sufficient in and of itself. Validity will probably be greater when the situation allows for personality to contribute to success on the criterion variable (e.g., Bem & Allen, 1974). For example, personality is more likely to make a difference in a manager's job as opposed to an assembly worker's job. In this vein, Barrick and Mount (1993) demonstrated that autonomy in the work setting is one situational characteristic that moderates the Conscientiousness-performance relationship. Specifically, they found that the relationship between job performance and three factors of the Big Five (i.e., Conscientiousness, Extraversion, and Agreeableness) increased as the level of autonomy in the job increased.

The findings of Barrick and Mount are consistent with the strong versus weak conceptualization of situations (e.g., Bem & Allen, 1974; Chatman, 1989; Weiss & Adler, 1984). Typically, weak situations are those in which there are few demands or pressures to conform. Under these conditions the person has a greater degree of latitude in determining behavioral choices. As such, individual differences in personality attributes are more likely to influence the behavior the individual adopts. Strong situations, on the other hand, are those in which there are considerable demands or pressures to conform. These situations restrict the range of behaviors that the individual is willing and able to exhibit. Therefore, one would expect Conscientiousness, and to a lesser extent other personality factors, to exhibit the greatest relationship in situations and with performance variables under the direct control of the employee.

Two such variables under the direct control of the employee include absenteeism and turnover (Martocchio, 1994; Martocchio & Judge, 1994; Nicholson & Payne, 1987). For example, Martocchio and Judge (1994) used policy-capturing to analyze individuals' decisions to be absent. They found a number of explanations for absenteeism, including hobbies or leisure activities, community or
religious activities, day of the week, kinship responsibilities, work demands, and personal illness.

Similarly, Nicholson and Payne (1987) found that individuals' choices to be absent differed even in the case of minor illnesses. These differences were attributed to the effects of absence norms that were thought to influence absenteeism rates (Martocchio, 1994).

Turnover is also thought to be an employee choice and is related to absenteeism (Porter & Steers, 1973). Mitra, Jenkins, Douglas, and Gupta (1992) in a recent meta-analysis, reported a corrected average correlation of .33 between absence and turnover. Although significant, the magnitude of this relationship suggests that the two variables do not form a unitary construct. Dalton and Todor (1993) suggested that absenteeism can be either a precursor to turnover or an alternative. In the precursor view, absenteeism is seen as a "pain-avoidance" strategy in which employees begin to absent themselves from the organization prior to leaving the organization permanently. In this case, the relationship between absenteeism and turnover would be positive. In the alternative "adjustment model," employees may find the job unsatisfactory, but as long as they are able to be absent, the job is bearable and they do not quit. This model suggests that high levels of absenteeism would actually lead to lower levels of turnover. Taken as a whole, both absenteeism and turnover are under the behavioral control of the individual and represent an autonomous situation in which Conscientiousness may influence action.

In sum, the importance of the situation in explaining behavior should be considered in conjunction with personality variables. Because personality will not completely account for all the variance in behavior, the situations in which personality will have the greatest influence should be further identified (e.g., autonomous contexts). Therefore, to maximize personality prediction, this study will focus on situations and two performance correlates under the control of the employee: absenteeism and turnover. Before exploring that possibility, however, it is first necessary to discuss some criticisms of personality measurement.
Personality Measures as Self-Report. Hogan (1986, 1991) has addressed the limitations of paper and pencil personality measures arguing that individuals' responses to self-report personality inventories are characteristic of the manner in which they present themselves throughout everyday life, including work-related settings. Responses to personality measures are not self-reports, but rather, reflect how individuals want to be regarded. According to Hogan, individuals' responses to particular items do not accurately reveal actual past behavioral tendencies. Instead, responses reflect behavior that they want others to believe they engage in, thereby influencing how others perceive them. Scale scores are hypothesized to reflect individuals' characteristic interpersonal style instead of underlying traits or tendencies.

For some (e.g., Kagan, 1988), however, paper and pencil tests are limited in their ability to adequately tap personality, and there continues to be concern regarding their ability to do so (see Briggs, 1992; Ozer & Reise, 1994 for reviews). Kagan contended that paper and pencil personality measures fail to capture the variability of personality across a range of situations and are influenced heavily by social desirability and mood. Support for Kagan's situational specificity assumption can be found in a study by Schmit and Ryan (1993). They examined the similarity of the factor structure of the Five Factor taxonomy between applicant and non-applicant samples and found that the Five Factor structure deviated within the applicant sample. They concluded that there may be a sixth factor in applicant samples that could best be described as an ideal employee dimension. They developed three arguments to explain their findings. The first, which is similar to the argument developed by Hogan (1986, 1991), suggests that job applicants may respond differently to items in order to produce a specific self-presentation. The second is that in an employment context, applicants may bring to memory a competent employee when responding to personality measures. Then, the applicant responds to items in a manner perceived to be characteristic of the known competent employee. Finally, the applicant may be responding to present an image consistent with an implicit idealized version of an employee. Regardless of the explanation, Schmit and Ryan argued that applicants may
be responding to particular items in dissimilar ways depending on their interpretation of the item with respect to performance at work. This, they argued, was particularly the case with many of the face-valid Conscientiousness items that were clearly related to work. Schmit and Ryan's conclusion that the factor structure of personality tests may change across populations is a question to consider when using personality measures. The argument proposed here is that rather than revising face valid items which have evidence of construct validity (Hough & Paullin, 1994), it may be more useful to implement alternate methods of scoring those items. The following section provides the theoretical rationale offered by interactional theory as a means of scoring existing items to take into consideration situational influences.

Interactional Approach. Theorists who adhere to the interactional approach maintain that individuals both choose and create the situations in which they participate (e.g., Schneider, 1983; Schneider, Goldstein, & Smith, 1995). As proponents of the interactional framework, Schneider and colleagues (Schneider, 1987, Schneider et al., 1995) maintained that personality is an important predictor of organizational performance. He posited that individuals choose situations consistent with their personalities and avoid situations that are inconsistent. Therefore, personality will be one characteristic that influences how people select themselves into and out of organizations. According to Schneider's Attraction-Selection-Attrition (ASA) framework, individuals who have obtained tenure, and are therefore successful performers, are apt to evolve into similar patterns of behavior over time. This similarity effect becomes even more pronounced due to selective attrition (Jackson, Brett, Sessa, Cooper, Julin, & Peyronnin, 1991), similar chance encounters, and experiences that influence individuals' personalities in the work context (Bandura, 1982). The net effect is that employees of an existing workforce should have similar responses to personality measures. Conversely, members who are different may be prone to withdrawal behaviors (Kristof, 1996), which include increased absenteeism (Johns & Nicholson, 1982) or turnover (Jackson et al., 1991).
Organizations as Situations. Research by Jordan, Herriott, and Chalmers (1991) offers support for the similarity of personality types within an organization. In their study, 344 United Kingdom managers from four organizations completed the 16 PF personality measure. Results showed that the personality types of managers were similar within organizations. The authors also found a significant interaction between occupation, organization, and personality such that for a given occupation there were different personality combinations present in the four organizations. Jordan and colleagues argued that this was an important discovery worthy of further research.

Continuing with this line of reasoning, B. Schneider and J. L. Schneider (1994) have taken the position that organizations can be statistically differentiated based on the personalities and the life histories of the people in them—even when the people are in the same occupation. Schneider et al. (1995) argued that organizations value different outcomes. This value structure, handed down by organizational leaders, results in individuals gravitating to organizations that are most congruent with their internalized values. For example, an accountant who values environmental issues may have objectives that are incongruent with those of an organization in the petrochemical industry. Were this particular accountant to accept employment within the petrochemical industry, the fit between the organization's and accountant's values may be so dissimilar as to cause the accountant to leave the organization or to exhibit other withdrawal behaviors (e.g., absenteeism; Schneider, 1987). Mael and Ashforth (1995) stated that identification with the organization is a critical component of organizational behavior, and individuals who do not identify are more likely to turnover or withdraw.

One outcome of individuals gravitating to organizations consistent with their values is that organizations may become increasingly homogenous due to attrition (Kristof, 1996; Schneider, 1987). This will be especially true in the case of the core values that are deemed important by the organization and the individual (Kristof, 1996; Mael & Ashforth, 1995). This is not to say, however, that there will be an absence of personality differences within organizations, since it is improbable that homogeneity of values will exist across all possible value dimensions (Kristof, 1996). Therefore,
whereas differences within organizations on values are present, the differences between organizations are likely to be greater (Schneider et al., 1995). Also, because personality and values are linked and reciprocal, there should be a similar relationship between personalities of organizations and individuals (Dawis, 1992; Kristof, 1996). Therefore, personality factors associated with core values will also become similar over time. Using the previous accountant example, in an organization that values environmental issues, Extraversion may be valued as employees are expected to champion environmental causes. However, in another organization being vocal and outspoken may not be of value, and Extraversion would explain little variance in performance. The end result is that when personality measures are aggregated across organizations and occupations, significant relationships between personality and performance may be obscured. Continuing with the example, although Extraversion was important in one instance it was not in another. The aggregation of the importance of Extraversion studied across organizations may lead to the conclusion that Extraversion is not an important characteristic when in some cases it is. Based on this reasoning, the use of personality measures across organizations may not be as useful an approach as identifying the important personality dimensions within an organization (Day & Silverman, 1989; Schneider et al., 1995).

Indirect support for the situational specificity of personality comes from the study by Barrick et al. (1993) who found divergent results from the findings obtained in the meta-analysis of Barrick and Mount (1991). The 1991 meta-analysis found a significant corrected correlation of .16 between Extraversion and sales performance. However, in the 1993 study, the relationship between Extraversion and sales performance was not significant (r = .03). Barrick et al. (1993) explained these discrepant findings as being a result of the type of sales job (e.g., retail or wholesale). They suggested that the retail industry requires a greater degree of Extraversion because individuals are required to interact with the public on a more frequent basis. On the other hand, in the wholesale industry, incumbents may be required to maintain existing accounts rather than to initiate new ones. However, an equally plausible explanation is that prescribed by the ASA model. It may be the case that retail
and wholesale industries attract and retain different personality types which vary on the Extraversion dimension.

Taken as a whole, the research presented above indicates that personality measures are best utilized to measure personality dimensions considered relevant to the organization. To achieve this end, personality dimensions should be identified that predict performance outcomes based on organization values that may differ at several levels including industries, organizations, and departments (Kristof, 1996). Therefore, it is important to consider the possibility that personality measures are situationally specific (Tett et al., 1991).

Test Administration as a Situation. Additional research suggests that not only may the industry, organization, and occupation influence the validity of personality measures, but so may the context in which the test is administered. For example, Schmit, Ryan, Stierwalt, and Powell (1995) assessed the differences in mean scale scores of the Big Five dimensions of personality and found that several of the scales differed depending on whether the test was administered under an "at work" context or a generic one (e.g., personality in a variety of contexts). Schmit et al. found that the criterion-related validity of the Conscientiousness scale of the NEO-Five Factor Inventory (Costa & McCrae, 1985) increased from .25 to .47 when items were reworded to reflect an "at work" context. However, the structure of the test and its psychometric properties remained the same despite alterations in item content and instructions.

Summary. Certain important personality constellations may differ within industries, organizations, and occupations. At a lower level of analysis, it appears that the context in which measures are administered may influence their validities (e.g., Block, 1995). Therefore, situational influences should be addressed with regard to paper and pencil personality measures. One method would be to study personality measures in a work specific context (e.g., applicant samples) and examine the person-situation interaction in this setting. Given that the factor structure of personality tests can differ between applicant and non-applicant samples (Schmit & Ryan, 1993), one might
assume that applicants are responding to personality items in an at “at-work” context; hence, differences in validity may be noted between the two samples (Barrett, Phillips, & Alexander, 1981).

In future sections, I discuss how biodata scoring may enhance predictive validity in both applicant and incumbent samples through the identification of items that best predict a given criterion (Mumford & Stokes, 1992). Also to be discussed is how using biodata scoring on existing personality measures may help bridge the gap between personality and biodata instruments. Indeed, Schmit et al. (1995) concluded that changing the content of the personality items in their study had the effect of making the questions more like biodata items than personality items. As such, it may be useful to blend the personality and biodata paradigms to increase their usefulness in selection contexts.

**Item Level of Analysis**

If personality researchers adopt the biographical scoring approach, the unit of analysis shifts towards the item level rather than the broader domains of the Big Five. Research at this level is important because theorists are in some disagreement regarding the specific facets that comprise the Big Five, especially at subordinate levels (e.g., Block, 1995). Indeed, several theorists have argued for further research at levels of analysis below the Big Five trait level because the description of personality at the five factor level is too broad (e.g., Block, 1995; Costa & McCrae, 1995; McCrae & John, 1992; Schneider & Hough, 1995). However, description of personality at subordinate levels below the Big Five has proven to be a difficult task (e.g., Block, 1995). Many contend that it is highly likely that personality factors at subordinate levels (e.g., domains or facets) interact with each other in complex ways that have yet to be established and hence pose difficulties for the researcher (e.g., Block, 1995; Ozer & Reise, 1994). Block (1971) maintained that analysis of personality scales should be conducted at an even lower level of analysis, claiming that studies at the item level can produce interesting findings.

Ozer and Reise (1994) provided a theoretical rationale for examining personality at the item level. They maintained that previous personality research has inadequately addressed the distinction between
latent and emergent variables. In a latent model the trait causes variation in its indicators (e.g.,
behavioral manifestations). In contrast, in an emergent model the trait is caused by its indicators. If a
latent model is assumed to be true, then the accepted standards for what constitutes a good measure of
a construct (e.g., high internal consistency) are applicable (Bollen & Lennox, 1991). However, Ozer
and Reise contended that most of the constructs of interest to personality researchers can be classified
as emergent. Consequently, this requires a different analytic framework in which the construct
validity of the measures is examined with an emphasis on external rather than internal criteria.
Further, an accurate assessment of emergent variables will examine not only simple additive
combinations of components but nonadditive, nonlinear combinations as well. Ozer and Reise cited
the example of leadership to argue their case. They contended that leadership is frequently
conceptualized as resulting from additive combinations of more basic variables (e.g., dominance and
extraversion) derived through a multiple regression framework. Although the resulting equation may
predict the dependent variable in situations where the independent variable is nonlinear or
nonadditive, it does not provide an explanation. The problem is even more complex in situations
where multifaceted constructs are used to explain behavior, such as personality. Therefore, the
approach adopted in this paper is to examine personality at the emergent level. Biodata offers one
technique of examining and expanding item level understanding of personality-criterion linkages (e.g.,
Russell, 1994).

Several studies have indicated that to effectively demonstrate the linkage between personality and
external criterion measures of work behavior, the relationship between the two must be based upon
sound theoretical rationale (e.g., Day & Silverman, 1989; Muchinsky, 1986; Tett et al., 1991).
Support for the importance of identifying a priori linkages comes from the Tett et al. (1991) meta-
analysis which found that studies using a confirmatory approach produced higher validities than those
using an exploratory one. Whereas the link between job performance and personality at the broad
levels has been demonstrated, there are fewer reported findings concerning the link subordinate levels
of personality and job performance (Hogan, 1991). One reason rests in researchers' and theorists' divergent conceptualizations of personality at the subordinate or facet level (Revelle, 1995). Another is the lack of independence between facets (Costa & McCrae, 1992; 1995). This in turn leads to a third problem in which single personality items are used to measure multiple facets (Hogan, 1991; Ozer & Reise, 1994).

Several theorists have stated that for personality research to achieve continued progress, advances must be made at the facet level (e.g., Costa & McCrae, 1995; Hogan, 1991; Ozer & Reise, 1994). In this vein, Costa and McCrae argued that facet analysis may be especially appropriate with regard to criterion-keyed scales (a technique commonly employed in the development of biographical keys). These authors stated that criterion-keyed scales are typically complex and multidimensional, measuring a combination of relevant characteristics rather than a unified psychological construct.

Using this foundation, the goal of the present study is to heed Costa and McCrae's suggestion and further understanding of personality at the facet level by using a technique successfully employed in biodata: empirical or criterion keying.

ADVANCES IN BIODATA

Notwithstanding recent theoretical advances with biodata, much of the success of the technique has been judged using an empirical standard. As documented in a comprehensive review of the biodata literature by Mumford and Stokes (1992), numerous studies have shown that biodata measures are effective predictors of a variety of criteria. Similarly, Rothstein, Schmidt, Erwin, Owens, and Sparks (1990) claimed that "Substantial evidence now indicates that the two most valid predictors of job performance are cognitive ability tests and biodata instruments" (p.175). Reilly and Chao (1982) and Hunter and Hunter (1984) compared personality measures to biographical instruments and found that the validity coefficients of the latter were far superior. However, these reviews should be interpreted with caution because they were conducted before the recent progress in personality research that has demonstrated marked improvements in predictive validity (Barrick & Mount, 1991; Tett et al., 1991).
Despite the reported differences in predictive validity between personality and biodata, some authors contend that both are designed to measure noncognitive correlates of performance (Mael & Hirsch, 1993; Mount et al., 1994).

In the past, the two fields of research took different theoretical paths to reach the same goal. For example, personality took the approach that one's internal states could predict behavior across situations. On the other hand, biodata were designed to tap historical events that were influential in determining one's behavior and identity using the premise that past behavior predicts future behavior (Mumford & Owens, 1987). Whereas this provided some guidance, biodata measures fell victim to the criticism that they were developed with little concern for the meaning of the items, or their hypothesized relationships with the criterion. Instead, a vast number of items were developed via the imagination of the researcher (Russell, Mattson, Devlin, & Atwater, 1990), and were retained based on an empirical relationship with the criterion (Nickels, 1994).

Researchers and theorists in the field of biodata have addressed this concern and are forging new theoretical ground (e.g., Mael, 1991; Russell et al., 1990). Yet, the introduction of theory into biodata has only served to underscore its overlap with personality rather than to magnify the disparity. In fact, Russell (1994) and Tenopyr (1994) have suggested that the integration of Big Five theory into biographical data is a fruitful area of study. Before commenting on that integration, however, it is necessary to lay some groundwork. First, I discuss recent advances in biodata, placing a special emphasis on its commonalities with personality research. Second, I will outline biodata scoring methodologies to (a) compare their effectiveness and (b) demonstrate how they might be integrated into personality research.

### Biodata Theory

Biodata theorists have responded to the cries of dustbowl empiricism and have developed three prominent theoretical paradigms to guide future research: the developmental-integrative model (Owens & Schoenfeldt, 1979), the ecology model (Mumford & Stokes, 1992), and social integration...
theory (Mael, 1991; Mael & Ashforth, 1995; Mael & Hirsch, 1993). The following section will provide a brief description of each approach and outline some of the commonalities with personality theories.

**Developmental-Integrative Model.** Owens and Schoenfeldt (1979) are considered to be among the pioneers of biodata theory and have provided empirical groundwork for subsequent theories. They demonstrated that historical events, captured through paper and pencil inventories, could classify individuals into subgroups. On the basis of individuals' initial subgroupings, Owens and Schoenfeldt were able to demonstrate in subsequent studies (cited in Owens & Schoenfeldt, 1979) significant differences in participants' future activities. Specifically, they conducted several studies examining subgroup differences in the following domains: interests, attitudes, values, and motives; cognition, creativity, and decision making; social processes; learning and memory; perception; physical-physiological; and personality. Owens and Schoenfeldt reported meaningful differences in 80% of the subsequent studies. With regard to personality, the authors reported differences between subgroups in 100% of their 20 studies. Their findings supported the hypotheses that (a) an individual is shaped by the environment and (b) one selects environments that are compatible with a self-percept and rejects others that are not. They stated that their findings contributed to the notion of the vital role of experience in the development of personality.

**Ecology Model.** The second major theory of biodata, labeled the ecology model (Mumford & Owens, 1987; Mumford & Stokes, 1992), shares common ground with the developmental-integrative model described above. The ecology model offers the following explanation of behavior: Initially, individuals are endowed with hereditary and environmental resources and restrictions that determine initial individual differences. Then, as they progress through life, people select situations based on the perceived reinforcement value of their outcomes. This perceived value is based partly on needs and values, and partly on preexisting characteristics. Adaptation continues through the identification, selection, and further refinement of the choices of situations that result in the attainment of desired
goals. This cycle of choice, development, and adaptation leads to a set of focused choices based upon one’s perceived needs and values. According to this theory, a successful person seeks out a variety of situations to satisfy all needs and values. Through this process the individual develops a cohesive pattern of choices. For biodata researchers, the issue becomes the identification and assessment of a range of previous behaviors and situations that predict subsequent choices and performance.

Social Identity Theory. Mael (1991) criticized the ecology model claiming that the theory focuses on volitional behaviors but does not take into account input variables. Briefly stated, input variables are things that are done to the person, and include all prior situational exposures, such as parental warmth, parental beliefs, and school and community characteristics (Owens & Schoenfeldt, 1979). Mael proposed social identity theory as a means of addressing this problem.

According to the tenets of this theory, each person has a self-concept that is composed of a personal and a social identity. Personal identity refers to attributes that are personal in nature and specific to the individual, such as bodily attributes and dispositional characteristics. Social identity includes those aspects of a person which help define the self as a result of one’s psychological membership to a perceived social category. For example, social categories such as nationality, political or social affiliation, or formal and informal groups help to provide social identity. One’s self-concept is derived through personal and social identity which then results in behaviors consistent with the self-definition.

The social identity framework for biographical development therefore posits that every experience or set of experiences that defines a person has the potential to shape subsequent behavioral patterns. As such, when a person associates with a team, club, school, or other psychological group, the person is believed to take on a common set of aspirations, preferences, values, and self-perceptions ascribed to by group members. In contrast to the ecology model, biodata items encompass not only choice-based adaptive responses of the individual, but also the effects of all characteristics internalized through identification with psychosocial groups.
In short, Mael's approach to biodata shares much in common with Bandura's (1982), Schneider's (1983), and Buss' (1989) approaches to personality. All of these authors have stated that a person's previous experiences, even those not occurring by choice, can influence subsequent behavior. For example, personality experiences that are not positive and involve failures or lost accomplishments may also influence behavior (Bandura, 1982; Buss, 1989). If biodata researchers accept Mael's argument that the adoption of characteristics of psychological groups thought to influence subsequent behavior are in many cases obtained through chance encounters, and conceivably in a broad spectrum of contexts, then the problem rests in the identification of historical events that may have been experienced by a diverse set of individuals (Sharf, 1994).

Commonalities with Personality Theories

As indicated in the previous discussion, many biodata theories include elements of personality in their description of behavior. The developmental-integrative model, the ecology model, and social identity theory, all addressed the importance of the interaction of situational variables and personality in determining behavior and subsequent behavioral patterns. To borrow Kanfer's (1991) terminology, a major difference appears to be that biodata theorists rely more on proximal indicators of behavior, whereas personality researchers rely on distal ones. In other words, the former are primarily interested in the antecedent behaviors that predict job performance, and the latter are concerned with the constructs that predict job performance as mediated by these antecedent behaviors. According to biodata theorists, many of the antecedents of job performance are related to situations or specific behavioral instances that predict subsequent performance (Mael, 1991; Mumford & Stokes, 1992). Consequently, to avoid adverse impact, biodata researchers must identify antecedents that are not related either directly or indirectly to demographic variables through socioeconomic ones (e.g., Kleiman & Faley, 1990; Sharf, 1994), or have not been differentially experienced by various racial or gender groups (Reilly & Chao, 1982).
Socioeconomic Correlates

One reason to focus on distal rather than proximal correlates arises when the item content of biodata instruments is associated with demographic or socioeconomic variables (Kleiman & Faley, 1990; Sharf, 1994). A distal correlate is defined as a personality construct that influences behavior and occupies a higher level in a hierarchy of prediction than does biodata (Tenopyr, 1994). Proximal correlates are behaviors that may be associated with performance and are at a lower level in the hierarchy. Further, proximal correlates may be influenced by a number of constructs at higher levels of the hierarchy. Whereas some proximal behaviors are associated with higher level constructs, they may also be associated with socioeconomic variables. For example, consider how demographic or socioeconomic experiences might influence responses to the proximal item “how many magazine subscriptions did your parents have when you were a child?” This item is thought to measure the construct of reading ability or education (Owens & Schoenfeldt, 1979), but may covertly result in adverse impact. Specifically, recent Census figures suggest a trend between race and income (Statistical Abstract of the United States, 1994). Therefore, in families with a smaller income it is likely that there are fewer resources available to subscribe to many magazines. Similarly, a greater percentage of minority group members were raised in single parent homes (Statistical Abstract of the United States, 1994). As such, the limited income of a single parent wage earner may not be used towards magazine subscriptions. In both situations, the applicant’s ability to read, or even enjoyment of reading, has not been assessed despite the hypothesized linkage with performance used in item development (Mumford & Owens, 1987). Conversely, a distal approach would focus on a construct such as Hogan and Hogan’s (1992) School Success and employ an item such as “I would rather read than exercise.”

Proximal Variables and Differential Experience

Developing proximal rather than distal items may also result in adverse impact due to differential experience. According to Reilly and Chao (1982), biodata researchers need to address this concern.
when developing items. Reilly and Chao and others (e.g., Kleiman & Faley, 1990; Sharf, 1994) reported that there are a number of behavioral experiences that may predict a specific work outcome. However, due to time constraints it is impractical to attempt to identify the complete set of experiences that may lead to similar outcomes. Therefore, the identified situations that remain in the scale based upon their predictive ability may reflect situations that have not been experienced by members of various protected groups (i.e., women and minorities). The following theoretical explanation may clarify how this situation might transpire. Both the Attraction-Selection-Attrition theory espoused by Schneider (1987) and the empirical work of Jackson and associates (1991) suggest that heterogeneity in the workgroup contributes to increased turnover, decreased instances of recruitment, and decreased promotion. As a result, the remaining members of the workgroup are likely to exhibit homogeneity across several characteristics, including personality. Because biodata instruments are typically developed using incumbents (e.g., Stokes, Hogan, & Snell, 1993), it may be the case that minority group incumbents used in the development of biodata instruments have come from backgrounds similar to majority group members. As such, using a sample of incumbents to develop and validate items may not be representative of the applicant population because differences in experiences in the incumbent pool may not be equivalent to those of applicants. A second explanation is provided by social identity theory. If one adopts the theory's assumptions (Mael, 1991) concerning group identity, then it is plausible to suggest that minority group members may not have adopted the self-defining characteristics subscribed to by the minority group, but rather those of the majority group. In other words, minority group members may have ascribed to the attitudes, values, and aspirations of the members of the organization, occupation, or work group (Kristof, 1996) rather than to those of the minority group.

Take for example a successful female firefighter. As a youth, she may have participated in a greater number of predominantly "male" activities. As her development progressed, she preferred these types of activities over predominantly "female" activities. Consequently, these preferences are
likely to be expressed when filling out a biographical questionnaire. Although she is similar to the group of firefighters (with whom she may identify according to Social Identity theory), her responses to the instrument may differ from the group of females. Hence when assessing item bias in incumbent samples, it may be difficult to detect group differences because some similarity of experiences is already likely to exist (Schneider et al., 1995). Using an incumbent sample to develop biodata instruments may result in unfair adverse impact when female applicants who possess many stereotypically "female" experiences, but are capable of performing the job, are not selected (Muchinsky, 1986; Sharf, 1994).

**Incumbent Samples and Item Development**

Given the above, the likelihood of discovering adverse impact in a concurrent sample is reduced because members in the organization may be similar in terms of personality (e.g., Schneider et al., 1995) or demographics (e.g., Jackson et al., 1991). However, to date, biodata has not been prone to claims of adverse impact (Muchinsky, 1986; Mumford & Stokes, 1992; Sharf, 1994). One reason may be that biographical instruments are commonly developed using concurrent samples (Mitchell, 1990; Stokes et al., 1993). As Mitchell (1990) indicated, biographical instruments are developed and validated in a tautological fashion such that similar behaviors are identified which operationalize both the predictor and the criterion. Further, this same sample may then be used to assess the measure's adverse impact even though minority and majority experiences and personality correlates are homogenous from the start. Continuing to use this instrument to select future applicants may further increase group homogeneity and decrease the organizations' long-term effectiveness (Schneider, 1987). Stokes et al. (1993) supported the differences between concurrent and predictive samples. They found that the empirical keys constructed for the incumbent pool differed significantly from those developed for the applicant pool. The authors attributed these differences to the effects of socially desirable responding in the applicant pool; although an equally plausible explanation is the similarity of incumbent characteristics explained by ASA theory.
This is not to say that using an incumbent pool is always detrimental. In fact, using an incumbent pool to develop an instrument may provide a benefit through the measurement of cultural variables. For example, if successful members in an organization subscribe to the culture of the organization (e.g., James & James, 1989) and membership in this culture explains variance in the criterion (e.g., James, 1982), as in the case of absenteeism (e.g., Martocchio, 1994; Mathieu & Kohler, 1990), then members may respond differently to personality or biodata instruments based upon their perceived membership with the culture. Moreover, because many biodata scale construction methods mandate that items that discriminate between groups be retained, it is likely that cultural variables also impact the items retained on the scale. And, because cultures are thought to vary from organization to organization (e.g., Kristof, 1996; Rousseau, 1985; 1988), it may prove useful to continue to develop biodata instruments or scale personality measures for a specific organization. In short, combining distal personality variables with biodata scoring may serve two purposes: (1) reduce differential experience that may be problematic with some biodata items, and (2) scale items that may contain variance influenced by culture or organizational context.

**Comparison at Different Levels**

**Theoretical Level.** The foregoing discussion suggests that to address the problem of identifying a set of experiences that are not correlated with protected group membership, it may be more effective for biodata researchers to focus on distal causes of behavior. By focusing on distal correlates found in the Big Five personality dimensions, it is likely that a larger set of individuals will exhibit similar characteristics than using a proximal measure, such as the number of part-time jobs held.

Though using different terminology, Tenopyr (1994) also advocated an integration of Big Five constructs into biodata research. She maintained that the two fields differ in their description of behavior at different levels of generality. Tenopyr posited that the two theories can be placed within a hierarchical model. In this model, the Big Five represent the highest level of generality. Because biodata focus on life history material at the behavioral and habit level, they assume a lower level in the
hierarchy. Much like the current stream in personality research (e.g., Costa & McCrae, 1995), Tenopyr maintained that constructs should be examined that fall beneath the Big Five level.

Mumford, Snell, and Reiter-Palmon (1994) went so far as to say "...that personality is closely related to, if not completely synonymous with, life history as measured by background data items" (p. 593). They based their conclusion on a review of studies that had examined the factor structure of biodata measures. One such review (Mumford & Owens, 1987) found factors frequently identified in background data literature such as adjustment, extraversion, cultural interests, social leadership, and social maturity that closely paralleled the Big Five dimensions.

**Item Development.** Along similar lines, Russell (1994) suggested that one fruitful avenue for selection research rests in the integration of personality theories with biodata item development. He asserted that biodata item construction represents a relatively untapped area but one that must be addressed to further the construct validity of biodata scales. Russell and others maintained that biodata items are typically derived through the researcher's imagination and most researchers fail to discuss how their items were derived (e.g., Mumford & Owens, 1987; Russell et al., 1990). Moreover, Russell (1994) contended that constructs in personality such as the Big Five offer an intriguing avenue for item development. He proposed that items could be developed that reflect antecedent events that produce differential response rates between upper and lower criterion groups using the Big Five personality dimensions. Consider the Extraversion dimension as an example; here, upper criterion group members may have experienced a greater number of situations that required them to be vocal, adventurous, or sociable.

**Item Characteristics.** The overlap between biodata and personality items, then, should be studied at both the item and the construct level (Hough & Paullin, 1994). As it stands now, many biographical theorists continue the practice of employing personality type items in their instruments (e.g., Russell et al., 1990). As such, it is oftentimes difficult to discriminate between the two types of items as indicated in the examples below, of which the first pair appear in the Hogan Prospective Employee.
Personality Inventory (Hogan, 1986) and the second pair are from the biographical instruments developed by Stokes et al. (1993) and Kluger and Colella (1993):

"In school I am or was usually in the upper part of my class"

"I frequently have indigestion"

"When working do you usually prefer to have many things going on at once?"

"I become sick when I see vomit"

The preceding examples demonstrate the similarity of item content with personality and biodata. Whereas, Mael (1991) agreed that there is substantial overlap between the two, he contended that this trend should not continue, and in response developed a classification scheme for biodata items. He and others (e.g. Asher, 1972) have argued that for an item to be classified as a true biographical item, it must be historical and verifiable. In support of this argument, some authors have reported that the use of verifiable items results in different responses than does using items that are unverifiable (Becker & Colquitt, 1992; Mumford & Owens, 1987). One reason may be that verifiable items are considered to be more resistant to faking than unverifiable ones and hence reduce attempts by applicants to portray a positive image (Shaffer, Saunders, & Owens, 1986).

Although the use of historical items has been advocated to distinguish biodata and personality items (e.g., Mael, 1991; Mael & Hirsch, 1993), there are few studies that have specifically examined how the time orientation of items influences validity. One exception is a study by Kleiman and Faley (1990) that addressed the question of how past- or future-oriented biodata items influenced validity. The authors developed a set of items that were congruent in content except for the specification of whether the behavior occurred in the past or in the present. Kleiman and Faley found that biodata items that measured present behavior had validity coefficients that were comparable, and in many cases superior, to an item set measuring past behavior. Consequently, the focus on past behaviors may not be the panacea for biodata researchers and may explain why many biographical instruments contain a mixture of the two items. Similarly, an issue that needs to be addressed by biodata
instrument developers is an operational definition of the term "historical." Does it refer to behaviors that happened a week ago, a month ago, in high school, or in some other time frame? Kleiman and Faley (1990) maintained that further investigation is warranted to determine the optimal historical time frame that produces the best prediction. Until the exact specification of a biodata item is concretely defined and adhered to by biodata researchers, biodata instruments are likely to continue to contain subjective, personality-type items.

Social Desirability. Although the use of objective, verifiable items is good in theory and is thought to reduce socially desirable responding (e.g., Lautenschlager, 1994), many researchers have found in practice that subjective items result in better validity (e.g., Hough et al., 1990; Mael & Hirsch, 1993). As indicated in the study by Stokes et al. (1993) and Hough et al., responding to items in a socially desirable manner may actually increase the validity of the instrument. Stokes et al. proposed a possible explanation using impression management literature, suggesting that applicants respond to biographical instruments in order to create favorable impressions. In turn, this may increase the predictive validity because for some occupations impression management skills may contribute to successful job performance. Along these lines, Ruch and Ruch (1967) demonstrated that the K scale of the MMPI, which is designed to predict response distortion, actually predicted performance of sales personnel.

In addition to impression management, responding in a socially desirable manner may be a proxy for measuring cognitive ability. To support this reasoning, Stokes et al. (1993) cited a study by Trent (1987) that demonstrated that cognitive ability is related to effectively responding to items in a socially desirable manner. In other words, when faced with a biodata item or personality item, individuals who have greater cognitive ability are better at discerning what is being asked, and therefore are more adept at determining the correct response. Consequently, responding in a socially desirable manner may increase prediction since it is an indirect measure of cognitive ability which has been shown to predict performance across occupations (e.g., Hunter & Hunter, 1984). Although responding to items in a
socially desirable manner may improve validity, some have attempted to reduce faking in biodata instruments by keying options, that is item alternatives, rather than items (e.g., Kluger, Reilly, & Russell, 1991). However, as the results from above suggest, attempts to reduce faking may not be necessary because dissimulation can contribute to validity. Instead, socially desirable items may be measuring latent variables such as impression management and cognitive ability. Along these lines, Buss (1989) and Hogan and Nicholson (1988) have suggested that responding in a socially desirable manner is in and of itself a personality characteristic.

**Response Similarities.** The preceding argument that individuals respond to biodata items to present a certain image is very similar to Hogan's (1986) assumption concerning personality measures. Hogan maintained that individuals respond to personality test items to portray a specific self-image. The findings gleaned from Stokes et al. (1993) also suggest that individuals may respond to biodata items to portray a specific image. With regard to the faking of biodata measures, Kluger and Colella (1993) found that one's capability to portray a self-image may occur regardless of whether the item is verifiable or unverifiable. Kluger and Colella reported that avoiding fakable items and implementing scoring techniques to mitigate faking may be unwarranted due to empirical benefits (i.e., higher validity). Taken as a whole, the evidence presented above suggests that the question of item characteristics (e.g., verifiable versus unverifiable) appears to lean in favor of the subjective item. Regardless of the item type, however, the more important question for biodata researchers remains in identifying and interpreting the latent traits associated with item responses (Russell, 1994).

Addressing this question may further research by examining biodata's relationship with personality. Similarly, exploring the latent traits associated with item responses may also increase understanding of personality at the facet level (Costa & McCrae, 1995; Ozer & Reise, 1994).

Whereas several theorists (e.g., Mumford et al., 1994; Russell, 1994; Tenopyr, 1994) have discussed the similarity between personality and biographical data, few studies have directly examined the relationship between the two. Mael and Hirsch (1993) attempted to fill this void. They developed
objective and verifiable biodata items that were empirically keyed to temperament scales and then
developed a priori scale values to predict performance on four criterion measures. Their goal was to
assess the overlap between temperament scales and biodata scales in their relationship to criteria.
Mael and Hirsch labeled their technique to develop the biodata scales as "quasi-rational." Their
approach was considered rational because items were not directly keyed to the criterion but rather to
temperament scales. Their approach was empirical because item values were determined by their
relationship to an external reference. Mael and Hirsch compared the five biographical scales that they
developed from the temperament scales through empirical keying to the actual temperament scales for
each of the four criteria. They found that out of these 20 comparisons (i.e., five scales by four criteria),
the biodata measures were superior in only 4 cases, smaller in 2 cases, and similar in the others.

Summary. The research reviewed above indicates that the content of the items may not be as
important as the theory used to derive them (Hough & Paullin, 1994). Further, because many biodata
items continue to employ a subjective format, it may be useful to heed contemporary thinking and
incorporate the Big Five into biodata (e.g., Russell, 1994; Tenopyr, 1994). The point of this paper,
however, is that the use of the Big Five as a framework in biographical item development may lead to
a number of items that are similar in content to existing Big Five measures. This being the case, it
may prove useful to apply the most effective biodata scaling techniques to personality measures
designed to measure the Big Five dimensions to maximize validity. In the next section, I review the
various approaches to biodata scaling to identify the method that may best achieve this goal.

Biographical Data Scaling Techniques

Scaling refers to the technique used to combine items and generate scoring of psychometric tests
(Murphy & Davidshofer, 1994). Similar to personality measures, there are essentially three types of
scaling used in biodata: rational, factorial, and empirical. In addition, Mael (1991) has advanced a
new strategy to biodata scaling referred to as rainforest empiricism. A brief definition of each method
is presented below.
**Rational Scaling.** The rational scaling approach specifies that a priori hypotheses be developed to achieve the goal of specified items grouping together into a scale. The emphasis on rational scaling is on the content and construct validity of the scaling system (Mumford & Owens, 1987). Fuentes, Sawyer, and Greener (1989) and Mumford and Owens (1987) reported that rational scaling procedures have not received much attention in the literature. Mumford and Owens suggested that future biodata research should examine the generalizability of rational scales to predict criteria over time, culture, and organizational settings. They also stated that, "rational scaling procedures would prove especially attractive whenever content and construct validity are salient concerns" (p. 17).

Mitchell and Klimoski (1982) provided one of the few examples of the success of the rational scaling technique to predict the criterion of license attainment in a sample of 698 prospective real estate sales associates. They used three sources to develop their rational scale: (a) an examination of the tasks performed by the associates, (b) the literature on career counseling and career development, and (c) the social and psychological theories describing the influence of background factors on career success. Then, Mitchell and Klimoski identified four a priori life history constructs: drive, conventional success, social orientation, and real estate involvement. They then developed items to measure these constructs, and contrary to empirical scaling, assigned items unit weights. Employing the rational scale in the cross-validation sample produced a correlation of .36 with license attainment. Although this correlation was smaller than that obtained with empirical keying, Mitchell and Klimoski concluded that the rational approach improved the stability and interpretability of the scale.

**Factorial Scales.** These scales are derived through the use of factor analytic techniques to identify psychologically meaningful dimensions. Unfortunately, the nature of biodata items has resulted in limited success for this procedure (Fuentes et al., 1989; Mumford & Owens, 1987). Specifically, items are typically generated to tap heterogenous constructs. Consequently, it is rare to find factor analytic solutions that account for a large proportion of total item variance (Mumford & Stokes, 1992).
Moreover, it is not uncommon for single items to exhibit loadings of .30 or above on one or more dimensions (Mumford & Owens, 1987).

**Empirical Scales.** The empirical approach is concerned with maximizing prediction rather than increasing conceptual understanding (Mumford & Stokes, 1992; Nickels, 1994). Empirical keying procedures are typically used to develop a set of item weights based on a statistical analysis of item-criterion relationships. Items are then selected and assigned weights based on their ability to differentiate criterion group performance (e.g., membership in upper and lower criterion groups). These weights comprise the scoring key that is then used to generate predicted scores on a background data measure. Any meaning ascribed to scores is generally limited to an assessment of predictive efficiency. Most empirical keying methods involve strategies for generating differential item weights using variance maximizing procedures developed to optimize criterion group discrimination. There are several keying procedures available to achieve this end, including the vertical percent, horizontal percent, correlation/regression, mean criterion, and rare response.

The vertical percent method involves the calculation of differences in response percentages for high and low criterion groups. The first step in the process is to identify high and low criterion groups based on individuals who are most and least successful on a chosen criterion measure. The second step involves the calculation of a percentage within each of the extreme criterion groups who selected each item alternative. Finally, a difference in percentage for each alternative is calculated by subtracting the percentage for the low criterion group from the percentage for the high criterion group. Within the vertical percent strategy, there are several methods of assigning weights. In the vertical 5% strategy a weight of 1 or -1 is assigned if there is a least a 5% difference between the high and low criterion groups. In the vertical 10% strategy unit weights are assigned only if the difference between high and low criterion groups is 10% or greater. In the vertical net weight strategy, Strong's (1926) tables are used to derive weights by transforming percentage point differences to "net weights."

Weighting is a joint function of the percentage of each criterion group endorsing an item alternative.
and the percentage point difference between criterion groups. Lastly, the vertical net unit weight strategy (England, 1971) transforms the net weights from Strong's tables into values that vary from 0 to 2. This transformation involves the magnitude of the net weights and the sign of the weights (e.g., positive or negative).

The horizontal percent method employs a similar three-step process as the vertical percent method. First, high and low criterion groups are established using individuals who are most and least successful. Second, the number of people in the high criterion group who endorsed an item alternative is divided by the total number of people in the high and low criterion groups who endorsed the alternative. Third, the resulting percentage is divided by 10 to obtain each weight.

The phi coefficient, correlation/regression, or pattern-scoring (Mumford & Stokes, 1992) methods involve investigating the relationships between each item alternative (response versus no response) and the criterion. Weights are then determined based upon the significance levels of the validities.

The mean criterion method (Devlin, Abrahams, & Edwards, 1992) does not require the use of extreme criterion groups. The weight for each item alternative is the mean criterion score for all of the respondents who chose that alternative. Subsequently, each respondent choosing that alternative is assigned the corresponding mean criterion score.

Finally, the rare response technique (Telenson, Alexander, & Barrett, 1983), unlike the other empirical keying methods, does not use item-alternative relationships with the criterion in selecting and weighting alternatives. Instead, the rare response method derives weights on applicants rather than on selectees who have criterion scores. Infrequently selected responses are assigned more weight than frequently selected responses. In addition, positive or negative signs are assigned to the weights based on the direction of their anticipated relationship with the criterion of interest.

Comparison of Scaling Techniques. Mumford and Stokes (1992) reviewed empirical work that has investigated the differences in weighting techniques. Their summary indicated that whereas percent scoring methods may yield higher validity coefficients, the regression method may exhibit less
shrinkage in cross-validation. However, in terms of describing differences between criterion groups, it was shown that a variation of pattern scoring, that is the discriminant function method, was superior to the horizontal percentage technique (Smith & George, 1987).

Conflicting results have been reported when comparing the percent and rare response methods. Aamodt and Pierce (1987) found the vertical percent method to be superior to the rare response method in each of their five samples. One caveat of their study is that they did not cross validate their findings. Therefore, it is unknown whether the superiority of vertical percent keys would have held up in a second sample. Telenson et al. (1983) found none of the three vertical percent or horizontal percent cross validities were significant, and only one of the three rare response cross validities was significant. However, the Telenson et al. study has been criticized because of the small criterion-based key-construction sample they used (Devlin et al., 1992).

As a means of addressing these conflicting findings, Devlin et al. compared several empirical keying procedures with respect to shrinkage and cross validation. They investigated the validity and cross validity of five different techniques (vertical percent, horizontal percent, phi coefficient, mean criterion, and rare response) using five different data sets (participants were Naval Academy applicants) ranging in number from 75 to 1,200. They found that although the mean criterion method had the highest validity in the construction sample, it failed to hold in the cross validation sample. The vertical percent difference and vertical net weight strategies were the two procedures with the highest cross validities. The correlation method (phi coefficient) was significant in only one sample and the rare response method failed to produce significance in any of the samples. Devlin et al. concluded that the vertical percent difference strategy is the most useful of the keying procedures.

However, within the vertical percent method, they concluded that the differentially weighted strategies (vertical percent difference and vertical net weight) were superior to simple unit-weighted strategies.

Further support for the use of empirical keys can be found in the study by Mitchell and Klimoski (1982) reported earlier. They compared the predictive power of empirical keys and factorial scales and
found that the former produced higher validities and cross validities than the latter. Nonetheless, the factorial scales exhibited less shrinkage in cross validation. They concluded that whereas factorial keys may exhibit greater stability and interpretability, they may not be as effective as empirical keys for prediction in a particular situation. As stated earlier, the empirical approach has been criticized for its failure to contribute to the understanding and construct validity of biodata scales. Even Guion (1965) has admitted that it would be a poor decision for a selection specialist to refuse to use something he or she did not fully understand, when it adds appreciably to the overall validity of a selection procedure. Therefore, one approach to increase understanding may be to integrate the benefits of construct and content validity inherent in the rational scale development with the increased predictive capability of empirical scoring.

Rainforest Empiricism. Taking this tack, Mael and associates (Mael, 1991; Mael & Ashforth, 1995; Mael & Hirsch, 1993) have proposed an alternative to the empirical approach termed "rainforest empiricism." This technique departs from the blind empirical method in the choice and keying of items and instead attempts to strike a balance between the rational and empirical methods. In Mael's approach, scales, which are created through a priori hypotheses, are empirically keyed to a given criterion. However, the keying is done rationally such that each item alternative is assigned a weight based on a logical analysis of its empirical relationship with the criterion. Therefore, items that may receive a weight in one of the percent strategies discussed earlier due to a sample specific nuance, might not receive a weight in the rainforest approach. Specifically, item alternatives producing similar relationships with the criterion may be assigned a single weight rather than their actual weights. For example, in the mean criterion strategy, item alternatives might receive the following weights for the item "How many days did you miss school in college?": None (2.8), one day (3.1), two days (3.4), three days (3.0), and four or more (3.7). In criterion keying a lower value would be assigned to three days than to one or two days although logically this would not make sense. Instead, Mael advocates grouping the response alternatives one day, two days, and three days, into one
category and assigning unit weights of 1 and -1 to alternatives none and four or more, respectively. His intent is to maximize stability and reduce shrinkage during cross validation. Much like the rational approach, this method allows for greater interpretability of instruments at the cost of predictive efficiency.

Summary. There has been some disagreement among biodata researchers as to the best method of scaling (e.g., Hogan, 1994; Nickels, 1994). Some have argued that scales should be developed in a rational manner and should avoid the blind empirical approach (e.g., Hough & Paullin, 1994). Others, like Devlin and colleagues (1992) have explored the various scoring methodologies to determine which method of scaling maximizes prediction which can hold up under cross validation. Finally, rain forest empiricism represents a new approach to scaling that combines the rational approach to scale development, with the empirical method of scoring. Based on the above findings, it would appear that integrating a good theory found in the Big Five with the best method of biographical scaling, the vertical percent method, should maximize prediction. The methodological approach to be taken in this study is similar to the rainforest approach with the notable exception that item weighting will be developed using relationships with external criteria. The cost of this method in terms of greater shrinkage in cross validation appear to be offset by greater predictive validity (Mitchell & Klimoski, 1982).

INTEGRATION OF THE THEORIES AND HYPOTHESES

The preceding sections outlined advances in personality theory, biodata theory, and biodata scoring techniques. From this review, it appears that personality and biodata theories share much in common, and in subsequent sections I specify how the integration of personality and biodata can benefit personnel selection. To achieve this goal, hypotheses will be presented that investigate personality correlates of absenteeism and turnover. The main focus of the study is to better predict these correlates by incorporating biodata scoring methods into existing personality measures. The effectiveness of this scoring methodology will be examined in three separate samples using several organizations to
provide a better assessment of the application of biodata scoring to personality measures in various contexts.

Importance of the Research

There are several benefits that can be gained from this research. First, recent testing theory has stressed the importance of studying tests at the item level of analysis (Drasgow, 1991). The methods used in biodata scoring achieve this end through the examination and identification of items that exhibit the greatest linkage with a given criterion measure (Mumford & Owens, 1987).

Second, items can be identified that maximize the distinction between individuals who are high on a criterion against those who are low on the criterion. In this manner, latent variables, including contextual or cultural factors, that contribute to differences in item responses by high and low criterion groups may be discovered through biodata scoring. This discovery may lead to a better understanding of the personality dimensions that are associated with a given culture (Kristof, 1996). An investigation at the item level of analysis serves two purposes: (1) it examines both the broad domains and specific facets of the hierarchical structure of personality (Costa & McCrae, 1995; Tenopyr, 1994); and (2) it analyzes the structure of personality through emergent rather than latent models (Ozer & Reise, 1994).

Third, because of cultural and contextual influences, it is expected that for some personality variables, in some organizations, a different combination of personality dimensions will predict criterion performance (Day & Silverman, 1989; Kristof, 1996). Further, because a different combination of items and personality facets may explain organizational behavior across organizations, industries, and even within occupations (Jordan et al., 1991; Kristof, 1996), scoring that is tailored for the particular context and culture may maximize validity. According to Buss (1989), situational influences, either direct or indirect, are often sampled in personality trait questionnaires. Therefore, by tailoring the scoring for specific situations this research will help to understand the interaction of personality and context in the explanation of behavior.
Fourth, this research can contribute to the accumulation of evidence concerning the role of the Big Five in explaining work behavior. Several theorists have stated that to further understanding of the Big Five dimensions, and to offer support for its construct validity (e.g., Hogan & Nicholson, 1988). Future studies should examine the relationship of these broad dimensions with external criteria (Ozer & Reise, 1994). Moreover, research should examine the dimensions both in isolation and as a unit to increase understanding of behavior (e.g., Block, 1995; Buss 1989).

Finally, this research can help solve the applied problem of maximizing the predictive validity of personality instruments within organizations to reduce the occurrence of absences and turnover. Because biodata scaling may increase validity within an organization, it is important to address the issue of generalizability. For this study, there are two issues of concern with respect to generalizability. One is whether the Big Five constructs can predict across organizations. A second is whether the technique of applying biodata scoring to personality measures can be generalized.

In the next section, hypotheses are advanced to specify how the technique of combining personality and biodata paradigms might benefit selection in applied settings. The hypotheses are grouped according to two major themes. One theme is the application of biodata scoring to existing personality instruments. These hypotheses address the relationship between personality constructs and absenteeism using biodata scoring principles. They will be tested using three samples to increase generalizability.

A second theme examines how Attraction-Selection-Attrition tenets impact the scoring and application of instruments through cultural and contextual influences. Hypotheses are developed that examine the influence of organizational personality on turnover. Because personalities may differ across organizations (e.g., Schneider, 1987), it is important to conduct studies with a variety of organizations. Therefore, in this study organizations are drawn from a school district, the public transit industry, and a large public agency.
Biographical Scoring of Personality Measures

The method of scoring to be used in this study closely parallels Mael's rainforest empiricism and quasi-rational technique of biodata instrument development (Mael, 1991; Mael & Ashforth, 1995; Mael & Hirsch, 1993). In accordance with Mael's approach, theory will guide item development. Rather than developing new biodata theories, however, I argue along with Tenopyr (1994) for the adoption of the Big Five as the theoretical foundation for scale construction. As indicated by the biodata theories reviewed in this paper, many biodata theorists have either directly or indirectly incorporated many aspects of the Big Five taxonomy into their approaches. Support for this contention is demonstrated in studies reporting similar factor structures of personality and biodata measures (e.g., Mumford & Owens, 1987) and the use of similar items. Therefore, whereas it is advantageous to continue with the investigation of situational and experiential correlates of the Big Five constructs through life history approaches (e.g., Mumford et al., 1994), an equally plausible approach, and a potentially useful one, is to use existing personality instruments scored through biodata methodology. Consequently, the linkage between Big Five constructs and job performance measures may produce higher validities when biodata scoring technology is used. Further, applying biodata scoring to existing measures offers a better understanding of the items and underlying traits that best differentiate between criterion group members (e.g., Russell, 1994).

The use of personality theory in personnel selection and assessment has already been discussed. As reported by Day and Silverman (1989) and others (e.g., Robertson, 1993; Tett et al., 1991), there is a place for personality measures when they are properly matched to a particular occupation and organization. However, Day and Silverman offered the caveat that there is nothing to indicate in advance which measures should be used in which situations and for what purpose. Accordingly, measures should be based on thorough job and organizational analyses and should be validated within the organization. The incorporation of biodata scoring (vertical percent method) will assist in the goal of improved prediction by developing empirical scales within a particular organization.
Despite the empirical nature of the scoring, items contained in the measures of this study conform to rational approach guidelines (e.g., Hogan & Hogan, 1989). The rational approach is based on the principle that items be developed to measure constructs (Hough & Paullin, 1994); which in this case are the Big Five personality dimensions. In contrast, items developed according to the empirical approach are designed primarily to predict external criteria (e.g., Hogan, 1994).

**Big Five and Absenteeism**

This investigation uses a combination of rational and empirical methodology to formulate hypotheses concerning the relationship between personality constructs and indices of performance not traditionally predicted by cognitive ability tests (e.g., England, 1971; Hogan, 1991). One such relationship supported by previous research is between Big Five constructs and counterproductive behavior as manifested by absenteeism (e.g., Hogan & Hogan, 1989; Murphy & Lee, 1994). Empirical support for this relationship has been demonstrated in several studies. First, Dunn et al. (1995) found that the Big Five dimensions of Emotional Stability, Conscientiousness, and Agreeableness were associated with counterproductive behavior. Second, Ones et al. (1993) found that integrity tests, that are designed to assess counterproductive behavior, primarily measure Conscientiousness, Agreeableness, and Emotional Stability. Finally, Murphy and Lee (1994) demonstrated that Conscientiousness, Adjustment (Emotional Stability), and Likability (Agreeableness) exhibited significant correlations with integrity tests (average \( r \) ranged between .44 and .33). Given this overlap, it is expected that individuals who score high on these dimensions will be absent less frequently.

Traditionally, the scoring of personality measures is done in a linear fashion (Hogan, 1991). Here, if respondents answer the item in the keyed direction their responses are given a unit weight. In contrast, biodata scoring allows for an examination of nonlinear relationships between personality test items and the criteria (Russell, 1994), and item weights may differ based on this relationship (England, 1971). Therefore, one would expect the relationship between personality factors and absenteeism to increase using empirical scoring as many personality variables are thought to be
nonlinear in nature (Schneider & Hough, 1995). For example, Schneider and Hough maintained that for some jobs, Agreeableness may conform to an inverted U-shape relationship with job performance because there is probably an optimal level of one’s ability to get along with others. Similarly, Extraversion may have an inverted logarithmic function with job performance because more of this personality dimension may not increase performance beyond some optimal level.

One benefit of biodata scoring is that items are assessed at a more specific level (Drasgow, 1991). In traditional personality scoring, all of the items designed to measure a specific construct contribute to that scale score. Analyses tend to concentrate on predicting broad domains of behavior with broad personality dimensions (e.g., Block, 1995). With biodata scoring, the focus of analysis shifts to the item level. Here, only items that discriminate between high and low criterion groups are included in the final scale, and therefore only a fraction of the original set of items might be included in the empirical scale. Further, if items do not discriminate, then biodata scoring is not likely to be effective because assigned weights are based on criterion group differences.

As such, biodata scoring does not ensure success. The success of the vertical percent method of scoring rests in items' ability to differentiate between high and low groups (England, 1971). Moreover, the effectiveness of an empirically derived scale rests in the scale's ability to predict in a cross validation sample (e.g., England, 1971; Mumford & Owens, 1987). For example, items that differentiate in a development sample may not predict in a cross validation sample due to idiosyncratic characteristics. Therefore, it is essential that the empirically-derived instrument be cross validated because some shrinkage is expected, and that cross validities serve as the measure of an empirically derived scale's success (Mumford & Owens, 1987). Hence the following hypothesis is proposed:

**Hypothesis 1:** A significant negative relationship will exist between Conscientiousness, Agreeableness, and Emotional Stability and measures of absenteeism and
these cross-validated relationships will be greater when using empirical scoring than when using traditional scoring.

Kluger et al. (1991) have shown that for biodata instruments, nonlinear keys explain variance above and beyond keys scored in linear fashion. As such, one would expect that personality keys scored in a nonlinear fashion would also provide incremental variance above and beyond keys scored in the traditional manner.

**Hypothesis 2:** Empirically keyed scales for Conscientiousness, Agreeableness, and Emotional Stability will provide incremental variance above and beyond keys scored in a linear manner.

Some have argued that personality traits do not operate in isolation, but rather function conjointly (e.g., Block, 1995; Ozer & Reise, 1994). Consequently, these authors recommend that traits be studied in combination rather than independently. Recently, Davison, Gasser, and Ding (1996) have proposed a method, termed Profile Analysis via Multidimensional Scaling (PAMS), to study combinations of individual difference variables.

Briefly stated, PAMS is a technique for deriving the most prominent profiles obtained from a battery of measures, or scales, in a given population. PAMS involves two steps: (1) an estimation of major profile patterns in the population, called prototypical profiles; and (2) an estimation of individuals' correspondence between their observed profiles and the prototypical profile of the population. Individuals' parameters can then be used to study the relationship between profile patterns and other variables, including absenteeism and turnover.

PAMS differs from other empirical approaches, such as cluster analysis, because PAMS first defines the prominent profiles and then determines how well each person resembles a given profile. Cluster analysis, on the other hand, analyzes pairs of persons to define groups or clusters of people with similar profiles. Once these clusters are obtained, a profile that best describes that cluster is
estimated. Therefore, a primary difference between the two approaches is PAMS is based on an explicit model of the test scores, whereas cluster analysis estimates the model from the clusters.

In the case of absenteeism, one would expect profiles of individuals who exhibit high levels of absenteeism to be characterized by low levels of Conscientiousness, Agreeableness, and Emotional Stability, and relatively high elevations for Openness to Experience and Extraversion. Conversely, individuals with low rates of absenteeism would have profiles characterized by high levels of Conscientiousness, Agreeableness, and Emotional Stability, and low elevations for Openness to Experience and Extraversion.

Hypothesis 3: The profile pattern for individuals with low rates of absenteeism will be characterized by high elevations for Conscientiousness, Agreeableness, and Emotional Stability and low elevations for Openness to Experience and Extraversion.

Because PAMS uses scale scores to derive profiles, the same technique could be applied to the personality scales developed through biodata scoring. Therefore, one would expect similar profiles between personality and biodata scoring. However, it is likely that the profiles developed from the latter will have more pronounced elevations because group differences are factored into the scoring and only items that maximize prediction are retained in the scale.

Hypothesis 4: The profile pattern derived from biodata scoring for individuals with low rates of absenteeism will contain high elevations for Conscientiousness, Agreeableness, and Emotional Stability, and low elevations for Openness to Experience and Extraversion.

As stated by Davison et al. (1996), each profile contains a parameter estimation based on a linear combination of the trait measures. Hypotheses 3 and 4 would therefore produce one level parameter for the personality scale and one for the biodata scale. According to Davison and Skay (1991) the
level parameter provides an overall characterization of one’s personality that is similar in concept to cognitive ability’s “g.” One method of testing the importance of each estimation would be to examine which of the two level parameters (i.e., personality or biodata) explains the most variance in the criterion (M.L. Davison, personal communication, April 12, 1996).

**Hypothesis 5:** The level parameter for the biodata scored personality measure will explain more variance in absenteeism than will the level parameter for the traditionally scored personality measure.

Davison et al. (1996) suggested that individuals’ correspondence weights also provide a method of studying the relationship of profiles to various criteria. For example, one might test the hypothesis that a profile is equally prominent in two groups (e.g., absent or not absent, terminated or employed, men or women). Differences between the two groups on a given profile can be tested by contrasting the mean correspondence weight for the first group to the mean for the second group (M.L. Davison, personal communication, April 30, 1996). Consequently, the following hypothesis is proposed:

**Hypothesis 6:** The mean correspondence weights will differ significantly for individuals with low absenteeism rates compared to individuals with high absenteeism rates.

**Conscientiousness and Absenteeism.** As indicated in prior meta-analyses, Conscientiousness appears to be the most important motivational construct in job performance (Barrick & Mount, 1991; Tett et al., 1991). Therefore, one would expect that (a) the empirically derived scale score for Conscientiousness will exhibit a greater relationship with counterproductive behavior than the other personality dimensions, and (b) the average weight assigned to Conscientiousness items will be greater than the weights for the other dimensions. The following hypotheses are proposed to test the prediction that Conscientiousness will occupy a greater portion of the predictor space than the other dimensions (Costa & McCrae, 1995).
Hypothesis 7: An empirical Conscientiousness scale will be a better predictor of absenteeism than will empirically derived scales for Emotional Stability or Agreeableness.

Hypothesis 8: Conscientiousness items will be assigned greater empirical weights than will the other items.

In sum, the hypotheses developed above are designed to test the relationships among the Big Five constructs and their relative importance in the prediction of absenteeism. The technique of applying biodata scoring to Big Five constructs was an integral component of the hypotheses to increase predictive validity.

Effects of Attraction-Selection-Attrition

Organizational Differences. Whereas research supports the need to key items to a particular criterion (e.g., Mumford & Owens, 1987; Russell et al., 1990), there is also evidence that suggests that specific-keying will be necessary for the organization and occupation as well (e.g., Day & Silverman, 1989; Jackson et al., 1989; Jordan et al., 1991). The findings of the Tett et al. (1991) meta-analysis suggest that personality dimensions may predict differentially for specific organizations. They found that in some situations the relationship between personality dimensions was positive, whereas in others it was negative. Consequently, keys developed specifically for an organization should produce higher criterion-related validity because they incorporate the effects of organizational values and culture (Schneider et al., 1995).

Jordan et al. (1991) suggested that within organizations there are significant differences in personality types that explain criterion performance. Similarly, Kristof (1996) proposed that organization values and individual personalities interact to determine who is successful in an organization. She maintained that individual personalities combine to form an organizational culture and that the proper fit of one's personality with the culture influences subsequent behavior. Further.
individuals who do not fit the culture or who hold values incongruent with the predominant culture are more likely to leave the organization (Schneider, 1987; Schneider et al., 1995). Consequently, empirical keying of personality measures will conceivably take an organization's culture into account and maximize prediction within an organization. This suggests that a different combination of items will contribute to scale scores in each organization.

Mael and Ashforth (1995) stated that one frequently adopts the values of the organization and both dispositional characteristics and life history experiences contribute to organizational identification. To test this assumption, they developed biodata scoring keys that successfully differentiated between those who left and those who stayed based on their identification with the organization. Mael and Ashforth examined the responses of 2,535 army recruits, and found four biodata factors that were able to predict recruits who remained: activities involving outdoor work or pastimes; a dependable, non-delinquent lifestyle reflecting socialization of institutional expectations; a general preference for group attachments; and diligent involvement in intellectual pastimes. The authors stated that factor one appeared to be organization specific and pertained only to military operations due to the outside nature of the job. However, the other three factors appeared to parallel Big Five constructs. For example, Mael and Ashforth's factor two was similar to the Conscientiousness dimension, factor three to the Agreeableness construct, and factor four to the Openness to Experience dimension. As such, one would expect these three constructs to contribute to turnover and that empirical scaling will maximize prediction by selecting and weighting items that best differentiate between criterion groups.

Barrick and Mount (1996) studied personality correlates of turnover and found that two of the five factors as measured by the Personal Characteristics Inventory were significant predictors. In two separate samples of long-haul semitruck drivers from different organizations, the authors reported corrected correlations between Conscientiousness and voluntary turnover of -.26 and -.26, and Emotional Stability and turnover of -.23 and -.21.
Although Mael and Ashforth (1995) and Barrick and Mount (1996) reached the same conclusion about the role of Conscientiousness in the turnover process; the studies differed with respect to the role of the other Big Five factors and turnover. The divergence in findings underscores the point made earlier regarding the importance of contextual factors (e.g., organization type or culture) in determining which combination of personality factors best predict turnover. Therefore, the following hypothesis is proposed to help resolve the differences in Mael and Ashforth’s and Barrick and Mount’s findings:

**Hypothesis 9:** Significant negative relationships will exist between Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability and turnover and these relationships will be greater using empirical, rather than traditional scoring.

Deciding to identify with an organization is inherently based on a number of multidimensional characteristics (Mael & Ashforth, 1995), and it is difficult to discern precisely what those characteristics might be because many may be emergent (Schneider & Hough, 1995). Consequently, the PAMS approach discussed earlier may delineate some of the characteristics that differentiate those who leave an organization from those that stay. This study’s approach is similar to Caldwell and O’Reilly’s (1990) Q-sort methodology such that profiles are developed to determine one’s person-organization fit and these profile fit scores are then correlated with subsequent performance measures. Because Caldwell and O’Reilly were able to demonstrate that the degree of one’s person-organization fit was an important correlate of job performance, it is similarly expected that the PAMS methodology of assessing person-organization fit could be used to compare the profiles of those who turnover to those who remain. Based on Mael and Ashforth’s (1995) and Barrick and Mount’s (1996) studies, it is proposed that these profiles will contain elevations for Conscientiousness, Agreeableness, Openness to
Experience, and Emotional Stability. The following hypotheses are designed to test the importance of the personality in person-organization fit in determining turnover.

**Hypothesis 10:** The profiles of those who remain in an organization will contain high elevations for Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability and low elevations for Extraversion.

Biodata scaling could also be derived for each of these scales to maximize prediction. By doing so, it is expected that the profiles would produce more dramatic divergence.

**Hypothesis 11:** The profile pattern derived from biodata scoring for individuals who remain in the organization will contain elevations for Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability and low elevations for Extraversion.

As a means of testing the effectiveness of the two scales, the parameter estimates for the traditionally scored and biodata scored scales will be compared. One would expect that the biodata scored scale would provide better prediction than the traditional scale, hence:

**Hypothesis 12:** The level parameter for the biodata scored personality measure will explain more variance in turnover than will the level parameter for the traditionally scored scale.

Hypotheses 9 through 12 described above are designed to test the impact of person-organization fit on turnover. These hypotheses were developed in accordance with person-organization fit literature (e.g., Chatman, 1989; Kristof, 1996) and are designed to test some of the theory's major assumptions with regard to individuals who leave the organization. It is expected that biodata scoring will assist in maximizing the prediction by scoring items that maximize group differences.
METHODS

OVERVIEW

To test the hypotheses of this study, analyses were conducted using three separate samples consisting of different organizations and occupations to improve the ability to generalize the findings, and to demonstrate the effectiveness of this technique in other contexts. A description of the samples, data collection, and criterion and predictor measures follows. Because the data were not collected for the purposes of the present research they are considered archival (Lane, Meibaum, & Schnuer, 1991). As such, the data were evaluated to assess their usefulness using the criteria recommended by Lane and associates.

SAMPLE ONE

Participants

Bus drivers from a large western school district (N = 472) participated in an initial study by Lin, Doyle, and Howard (1990). Sample characteristics were 50.7% female (463 reported their gender); 78.5% were between the ages of 18 and 45 (461 reported their ages); 53.7% were African American, 22% were White, 20.5% were Hispanic, and 3.8% reported "other" (463 responded). The average length of tenure with the Transportation Department was 6.5 years. Two-thirds of the sample was used to develop empirical keys. The remaining third was used as the cross-validation sample.

Data Collection

Data were collected in 1989 for a previous unpublished validation study. Participants were allowed to complete measures during work hours and received their hourly wage while completing the measures. Several performance measures were collected for the initial validation study; however, for the purpose of the present study, the focus was restricted to measures of absenteeism and turnover. This information was extracted from the organization's payroll database.
Absenteeism

One criterion measure for this study consisted of absenteeism data. A job analysis indicated that absenteeism presented a significant problem for the organization due to scheduling requirements and the difficulty in covering vacant shifts. For this sample, absenteeism was operationalized as the total number of days away from work during the past year (Johns, 1994a). This information was obtained from each participant’s payroll records.

Turnover

Turnover data for this study were obtained six years after the administration of the personality test from the principle author of the aforementioned validation study. Approximately 25% of the original sample had quit and 8% had accepted promotions. The organization did not maintain records on the reason for separation and hence voluntary and involuntary turnover could not be differentiated in this sample.

Personality Measure

Personality was measured by the Prospective Employee Personality Inventory (PEPI; Hogan & Hogan, 1992). This measure was developed using the Big Five personality framework and is a derivative of the Hogan Personality Inventory (HPI; Hogan, 1986). The PEPI is a 198-item measure containing scales (the Five-Factor counterparts in parentheses) for Intellectance (Openness to Experience), Adjustment (Emotional Stability), Prudence (Conscientiousness), Ambition (Extraversion), Sociability (Extraversion), and Likability (Agreeableness). It should be noted that Hogan has divided the Extraversion dimension into Ambition and Sociability factors (Lifton & Nannis, 1990). His rationale is that whereas both components are related to Extraversion, for certain applications it is useful to distinguish between the "ascendancy" and "status seeking" components of Ambition from the "social interest" and "enjoyment of social interaction" aspects of Sociability. He suggested that the two components of Extraversion are not complementary in that Ambition is most related to the ethic of hard work (i.e., Conscientiousness), whereas Sociability is the dimension of
Extraversion most related to warmth and affiliation (i.e., Agreeableness). The PEPI used in this sample differs in many respects from its current form (Hogan & Hogan, 1992) and many of the items used in this sample were considered exploratory (Lin et al., 1990) and removed in future revisions. Cronbach’s alphas for the six scales in the Lin et al. study were as follows: Adjustment (α = .83), Ambition (α = .79), Likeability (α = .79), Intellectance (α = .75), Prudence (α = .67), and Sociability (α = .65).

**Empirical Scoring**

England’s (1971) vertical percent method was used to determine item weights. For each item, the relative frequencies of responses for the high criterion group were contrasted against the low criterion group. For example, for the question “I am sensitive to other people’s moods,” 90.7% of members in the high absenteeism criterion group responded affirmatively to the item compared to 80.4% of the members of the low criterion group. The differences between the cumulative frequencies of the groups’ responses were then calculated resulting in a net difference for each item option. In the example indicated above, there was a difference of 10.3% between the groups. From here, a net weight for each item option was assigned using Strong’s tables (1926; reported in England, 1971) that are presented in the Appendix. The net weight is a function of the percentage of each criterion group endorsing an item alternative and the percentage difference between the criterion groups. Strong’s tables contain three columns. Column One, is used when both percents (i.e., 90.7% and 80.4%) are between 8 and 92. When the original percents fall within this range, a difference of 10.3 results in a net weight of “2.” Column Two is used when one percent is between 3 and 7, or 93 and 97. A difference of 10.3 percent, when one of the percents falls within this range, would result in a net weight of “3.” Finally, Column Three is used when the percentage of criterion group members responding to an item alternative is between 0 and 2, or 98 and 100. Here, a difference of 10.3 percent results in a net weight of “4.” Using the original example reported above, Column One was used.
because both percents (i.e., 90.7% and 80.4%) were within the range of 8 and 92. As a result, a net weight of “2” was assigned.

Once the net weights were determined for each item, scores were computed for each of the PEPI primary scales by summing the empirically weighted items contributing to that scale. In addition, respondents’ scores were summed to compute a total score.

SAMPLE TWO

Sample Two was used to assess the utility of the biographical scoring technique using different organizations. Data for this study were obtained from a number of organizations to address the issue of personality combinations differing in various organizations (Schneider et al., 1995). The multiple organizations of Sample Two provided a means of constructing an empirical scale using similar, though not equivalent, organizations within the same industry. Because turnover data were not available for this sample, absenteeism served as the only criterion measure. Therefore, only Hypotheses 1 through 8 could be tested.

Participants

Data for the second sample were previously collected for the development of a selection battery for coach operators employed in the transit industry (Jacobs, Conte, Day, Silva, & Harris, 1996) and were obtained for this study from the principle author. The sample contained 76.0% males, 34.6% Caucasians, 10.6% Hispanics, 48.3% African-Americans, 2.2% Asian, and 1.7% who did not report. The sample characteristics are presented in Table 1. To have an adequate sample size to develop scales (England, 1971), the data for the agencies were combined to form one sample. As with Sample One, two-thirds of participants comprised the developmental sample and the remaining one-third comprised the cross-validation sample.
Data Collection

Data were collected from ten transit agencies located throughout the United States. Participants voluntarily completed the measures during work hours and were paid their hourly wage. The number of employees from each organization is also provided in Table 1.

Table 1. Sample Two Characteristics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Austin</th>
<th>Chicago</th>
<th>Dallas</th>
<th>Flint</th>
<th>Miami</th>
<th>N.Y.</th>
<th>Others</th>
<th>Seattle</th>
<th>State College</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N</td>
<td>84</td>
<td>49</td>
<td>100</td>
<td>217</td>
<td>88</td>
<td>120</td>
<td>94</td>
<td>94</td>
<td>18</td>
<td>864</td>
</tr>
<tr>
<td>Males</td>
<td>66</td>
<td>35</td>
<td>92</td>
<td>140</td>
<td>71</td>
<td>99</td>
<td>76</td>
<td>62</td>
<td>16</td>
<td>657</td>
</tr>
<tr>
<td>Females</td>
<td>18</td>
<td>14</td>
<td>8</td>
<td>77</td>
<td>17</td>
<td>21</td>
<td>18</td>
<td>32</td>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td>White</td>
<td>29</td>
<td>5</td>
<td>12</td>
<td>78</td>
<td>7</td>
<td>37</td>
<td>71</td>
<td>50</td>
<td>11</td>
<td>299</td>
</tr>
<tr>
<td>Hispanic</td>
<td>29</td>
<td>4</td>
<td>14</td>
<td>2</td>
<td>25</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Black</td>
<td>27</td>
<td>30</td>
<td>73</td>
<td>117</td>
<td>56</td>
<td>68</td>
<td>4</td>
<td>32</td>
<td>1</td>
<td>417</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>6</td>
<td>15</td>
</tr>
</tbody>
</table>

Absenteeism

The criterion measure for this study consisted of absenteeism data. A job analysis indicated that absenteeism presented a significant problem for the organizations due to scheduling requirements and the difficulty in covering vacant shifts (Coste, Tesluk, & Jacobs, 1996). Similar to Sample One, absenteeism was operationalized as the total number of days away from work during the past year (Johns, 1994a). The data were extracted from each organization's payroll records by the principal authors of the validation study.

Personality Measure

The personality measure used in this sample was the Hogan Personality Inventory (Hogan & Hogan, 1992). The HPI contains many of the items used in the PEPI. The benefit of the HPI rests in a larger pool of studies assessing its validity and a larger pool of items (Hogan & Hogan, 1992). As with the PEPI, this instrument is designed to measure the Big Five constructs. The HPI was developed using a rational or deductive approach and consists of 206 items developed to assess six components of personality: Intellectance, Adjustment, Prudence, Ambition, Sociability, and Likability. Similar to the
PEPI, Hogan has divided the Extraversion dimension into an Ambition factor containing ascendancy and status seeking components, and a Sociability factor containing social interest and enjoyment of social interaction components (Lifton & Nannis, 1990). One of Hogan's chief objectives with the HPI and the PEPI has been to demonstrate that personality measures, when properly developed and administered, can predict work-related performance criteria. Cronbach's alphas computed for this sample were .72 for Likeability, .82 for Sociability, .83 for Intellectance, .83 for Prudence, .87 for Ambition, and .89 for Adjustment.

**Empirical Scaling**

Empirical keys for the HPI were developed using the vertical percent methodology described in Study One. Empirical keys were developed by aggregating the data from the ten organizations and splitting this sample into a two-thirds developmental group and a one-third cross-validation group. The response frequencies were compared and items were weighted according to the magnitude of the group's differences.

**SAMPLE THREE**

Sample Three used archival data to assess the effectiveness of biodata scoring using an applicant sample rather than an incumbent sample. As indicated by Schmit and Ryan (1993), because of the differences in factor structure between applicant and non-applicant populations, it is important to study personality measures in contexts where individuals are completing the forms in what they perceive as an "at-work" context. Data were collected from individuals applying from January, 1991 to December, 1995 for the sworn peace officer positions of sheriff, marshal, and probation officer; investigative personnel assigned to the courts; and safety sensitive classifications assigned to administration. The collection of criterion measures was performed in a predictive rather than concurrent manner, which as Barrett et al. (1981) have indicated, may reduce the effects of range restriction.
An additional advantage of this sample is that the keys could be developed for two occupational groups (Schneider & Schneider, 1994). Although the broad occupational group of peace officer is similar and the law enforcement officers represent the same public agency, these officers rarely interact with one another and each department has specific functions. Furthermore, each law enforcement group has different policies, procedures, and rules because of different labor union representation. The primary purpose of using various occupational groups is to increase the generalizability of the findings rather than for statistical comparisons.

Participants

Data were collected from applicants for law enforcement agencies and other safety sensitive positions. Applicants for these positions are required by law to complete a psychological screening prior to employment. Only applicants who were hired between the dates of January 1, 1991 through December 31, 1995 were included in this sample. The total sample size across occupations was 1841, and the number of individuals used for each analyses below differed depending upon the employee’s date of hire. Of the total sample, 413 (22.4%) were female and 1428 (77.6%) were male. The average age was 29.12 with a standard deviation of 7.56. The ethnic composition as defined by Census categories was (percentage in parentheses): Blacks (11.9%), Hispanics (18.7%), Asians/Pacific Islanders (4.3%), American Indian/Native Americans (0.9%), Whites (60.3%), and Filipinos (3.5%).

Absenteeism

One criterion measure used for this study was absenteeism data. Absenteeism was operationalized as the total number of sick leave hours used during the past year (Johns, 1994a). Unlike the previous two samples, the data were not reported by days because employees of this agency are allowed to take sick leave as needed. The data were maintained in the employees’ payroll records as the number of hours taken. Absenteeism data covered the period of January, 1991 to June, 1996. Analyses for the hypotheses are presented separately for each year.
Turnover

Turnover data for this study were obtained for the years 1991 through 1996. The employee's reasons for leaving the organization were used in this study to differentiate between voluntary and involuntary terminations. Only reasons coded as “voluntary” were included in the subsequent analyses. Of the original sample, 532 (26.3%) terminated employment between 1991 through 1996.

Personality Measure

The personality measure used for this sample was the California Psychological Inventory (CPI: Gough, 1987). The CPI contains 462 true-false items designed to measure various facets of the normal personality. Items are grouped into 20 scales designed to measure attributes of personality involved in interpersonal behavior and social interaction. Although not developed specifically for measurement of the Big Five constructs, McCrae, Costa, and Piedmont (1993) have studied the CPI and determined that it is correlated with Extraversion, Openness to Experience, Agreeableness, Emotional Stability, and Conscientiousness as measured by the NEO-Five Factor Inventory (Costa & McCrae, 1985). To provide further support for the CPI as a measure of the Big Five, McCrae et al. (1993) rated the item content of the CPI scales in terms of the five factors and found adequate representation of the factors. Consequently, McCrae et al.’s ratings were used to assign items to the Emotional Stability, Extraversion, Conscientiousness, Agreeableness, and Openness to Experience dimensions. Cronbach’s alphas computed for the five scales using the entire sample were .87 for Emotional Stability (Neuroticism), .81 for Extraversion, .82 for Conscientiousness, .73 for Agreeableness, and .69 for Openness to Experience.

Empirical Scales

Empirical scales were developed in the same manner as in Sample One and Two. Only the Sheriff and Probation occupational groups had sufficient numbers to develop empirical scales for the turnover comparisons. An empirical scale was developed using an aggregate sample that included other employees that are required to take the CPI as a condition of employment. Two-thirds of each
sample were used to develop empirical keys. The remaining third were used as the cross-validation sample.
RESULTS

SAMPLE ONE

Scale Development

Two-thirds (n=315) of the original sample of 472 bus drivers were used to develop the scales for absenteeism and turnover. The remaining third were used as the hold-out sample for cross validation purposes.

Mumford and Owens (1987) maintained that item differences may sometimes occur as a result of sample nuances. Therefore, the hold-out and development groups were tested for differences in demographic variables and mean differences for the six scales of the PEPI. The results of the Chi-Square analyses presented in Table 2 and the univariate ANOVAs, presented along with the means and standard deviations, in Table 3 indicate that there were no significant differences between groups.

Table 2. Sample One Demographic Characteristics.

<table>
<thead>
<tr>
<th>Sample</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Other</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hold-Out (n=157)</td>
<td>22.3%</td>
<td>54.3%</td>
<td>20.3%</td>
<td>1.0%</td>
<td>3.2%</td>
<td>51.1%</td>
<td>48.9%</td>
</tr>
<tr>
<td>Development (n=315)</td>
<td>22.3%</td>
<td>56.1%</td>
<td>17.8%</td>
<td>0.6%</td>
<td>3.2%</td>
<td>48.4%</td>
<td>51.6%</td>
</tr>
<tr>
<td>Chi-Square Results</td>
<td>$\chi^2(4, 472) = .56, p = .97$</td>
<td>$\chi^2(1, 472) = .01, p = .90$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Mean Differences for PEPI Scales.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Adjustment</th>
<th>Ambition</th>
<th>Likeability</th>
<th>Intellectance</th>
<th>Prudence</th>
<th>Sociability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop (n=315)</td>
<td>$X = 76.82$, s.d. = 5.93</td>
<td>$X = 81.18$, s.d. = 4.07</td>
<td>$X = 39.87$, s.d. = 4.53</td>
<td>$X = 56.12$, s.d. = 5.02</td>
<td>$X = 46.50$, s.d. = 3.90</td>
<td>$X = 56.06$, s.d. = 3.97</td>
</tr>
<tr>
<td>Hold-Out (n=157)</td>
<td>$X = 75.99$, s.d. = 6.83</td>
<td>$X = 81.19$, s.d. = 4.07</td>
<td>$X = 39.97$, s.d. = 4.73</td>
<td>$X = 56.28$, s.d. = 5.24</td>
<td>$X = 46.94$, s.d. = 4.54</td>
<td>$X = 56.34$, s.d. = 3.83</td>
</tr>
<tr>
<td>F-Tests (1,470)</td>
<td>1.82, p = .18</td>
<td>.01, p = .97</td>
<td>.05, p = .83</td>
<td>.10, p = .75</td>
<td>1.24, p = .27</td>
<td>.55, p = .46</td>
</tr>
</tbody>
</table>

In addition, no significant differences were noted between the hold-out ($X = 5.05$, s.d. = 6.46) and developmental groups ($X = 4.96$, s.d. = 6.01) for the absenteeism measure [$F(1, 470) = .03, p = .87$].

Scales were developed separately for the absenteeism and turnover criterion measures. For the absenteeism scale, the item options of employees with high rates of absenteeism were compared to
those with low rates of absenteeism. This required splitting the continuous absenteeism variable into three equal groups and using the extreme groups for scale development. Whereas Mumford and Owens (1987) and Russell et al. (1990) have advocated the use of this method, Russell (1994) stated that there is not much guidance in the biodata literature as to the appropriate split, and there exists a great deal of researcher discretion in selecting a cutoff value. As such, the present study used the upper-third, lower-third used by Russell et al. (1990) as a guide. The low absenteeism group consisted of individuals who were absent less than 1.63 days for the year, whereas the high absenteeism group consisted of employees who were absent more than 5.40 days for the year. As stated by Linn et al. (1990) in their initial validation study, absenteeism is a pervasive problem in the transit industry and over five days of absences is considered problematic. From here, the response frequencies of the high absenteeism group were compared to the low absenteeism group for each item of the PEPI. The differences in frequencies were compared and assigned weights using Strong's (1926) tables. It must be noted that the response frequencies of the absenteeism groups did not differ for many of the PEPI items and hence these items did not contribute to the final scale score. Specifically, only 66.7% (24 of 36) of the Adjustment, 41.8% (15 of 35) of the Ambition, 58.3% (14 of 24) of the Likeability, 66.7% (24 of 36) of the Intellectance, 67.6% (23 of 34) of the Prudence, and 60.6% (20 of 33) of the Sociability items were used to develop the biodata scored scales. In addition, there was a reduction in the reliability of all six scales (α = .02 for Adjustment, α = .12 for Ambition, α = .01 for Likeability, α = .29 for Intellectance, α = .15 for Prudence, and α = .08 for Sociability) suggesting that the biodata scored scales are measuring multi-dimensional constructs (Mumford & Owens, 1987). Consequently, a total score was derived by summing the items that best predicted the criterion.

The turnover scale was developed by comparing the response frequencies of those who left the organization to those that remained. The ethnic composition of the turnover group (n = 143) was 0% Asian, 57.3% Black, 14.7% Hispanic, 25.9% White, and 2.1% that did not report. The ethnic composition of those who remained in the organization (n = 329) was 1.2% Asian, 53.8% Black.
21.6% Hispanic, 19.8% White, and 3.6% who did not report. The turnover and incumbent groups did not significantly differ \[ \chi^2 (4, N = 472) = 6.88, p = .14 \]. Similarly, the gender of the turnover (53.1% Female and 46.9% Male) and incumbent (48.9% Female and 51.1% Male) groups did not significantly differ \[ \chi^2 (1, N = 472) = .71, p = .40 \].

The development and hold-out samples were comprised of the same participants as the absenteeism scale. Potential sample nuances were tested for turnover rates using a chi-square analysis that showed similar rates for the hold-out (30.5%) and development (29.9%) groups \[ \chi^2 (1, N = 472) = .014, p=.90 \]. A total scale score was also developed for the turnover group to maximize prediction.

The percentage of items included in the biodata scored scales were 55.6% (20 of 36) of the Adjustment items, 72.1% (25 of 36) of the Ambition items, 66.7% (16 of 24) of the Likeability items, 58.3% (21 of 36) of the Intellectance items, 55.9% (19 of 34) of the Prudence items, and 57.6% (19 of 33) of the Sociability items. The reliabilities of the biodata scored scales as measured by Cronbach’s alpha were .05 for the Adjustment scale, .23 for the Ambition scale, .37 for the Likeability scale, .31 for the Intellectance scale, .17 for the Prudence scale, and .01 for the Sociability scale. These findings are not surprising in light of Mumford and Owens' (1987) assertion that biodata scales measure multidimensional constructs and hence the measurement of reliability through internal consistency approaches may be inappropriate.

Test of Hypotheses

Hypothesis 1 posited that the relationship between Conscientiousness, Agreeableness, and Emotional Stability would be greater using empirical scoring rather than traditional scoring. In order to control for the number of tests being conducted a conservative level of .008 (.05/6 = .008) was used to assess the significance of the correlations (Cohen & Cohen, 1983). Using this standard none of the personality measures were significant for the development group as shown in Table 4. For the hold-out group, only Openness to Experience (Intellectance) was significant as a predictor of absenteeism (r
As shown in Table 4, five of the seven empirical scales were significantly correlated for the development group, but the scales were not significant for the cross validation group. Nonetheless, Sociability (Extraversion, $r = -.17$, $p = .018$) and the Total scale score ($r = -.18$, $p = .012$) were significant at the conventional level of .05, and Emotional Stability (Adjustment, $r = -.12$, $p = .072$) and Conscientiousness (Prudence, $r = -.12$, $p = .064$) produced weak correlations (Cohen, 1988).

Although the correlations are not statistically significant, their magnitude is similar to the uncorrected correlations reported by Barrick and Mount (1991), and Tett et al. (1991) between personality and various criterion measures.

Hypothesis 1 also posited that the relationship between Conscientiousness, Agreeableness, and Emotional Stability would be greater using empirical scoring than traditional scoring. Meng, Rosenthal, and Rubin (1992) reported that comparing correlated correlation coefficients using the Fisher Z transformation is more appropriate than Hotelling’s t-test and was therefore used as the method of comparison in this study. Meng et al. (1992) argued that their approach offers a greater
degree of precision and the added advantage of creating confidence intervals to compare correlated correlations.

Although Conscientiousness (Prudence), Agreeableness (Likeability), and Emotional Stability (Adjustment) were the primary focus of the study, all six pairs of cross validated correlations were tested for significance. The results of the Z (normal curve) test and the 95% confidence intervals are presented in Table 5. As shown, the personality \( (r = .12) \) and empirical \( (r = -.17) \) correlations for

<table>
<thead>
<tr>
<th>Scale Compared</th>
<th>Z Score</th>
<th>Significance</th>
<th>Upper Con. Int.</th>
<th>Lower Con. Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment</td>
<td>.22</td>
<td>p = .587</td>
<td>.252</td>
<td>-.200</td>
</tr>
<tr>
<td>Ambition</td>
<td>-.73</td>
<td>p = .233</td>
<td>.109</td>
<td>-.239</td>
</tr>
<tr>
<td>Likeability</td>
<td>-.99</td>
<td>p = .161</td>
<td>.089</td>
<td>-.269</td>
</tr>
<tr>
<td>Intellectance</td>
<td>1.36</td>
<td>p = .086</td>
<td>.257</td>
<td>-.046</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.29</td>
<td>p = .387</td>
<td>.195</td>
<td>-.262</td>
</tr>
<tr>
<td>Sociability</td>
<td>-2.49</td>
<td>p = .006</td>
<td>-.061</td>
<td>-.513</td>
</tr>
</tbody>
</table>

Note: \( n = 157 \)

Sociability (Extraversion) were the only correlations that were significantly different \( (Z = -2.49, p = .006) \) due to a difference in sign. When the correlations were tested using absolute values, the correlations did not differ significantly \( (Z = .43, p = .33) \). In sum, the hypothesis that biodata scoring would significantly improve the correlations was not supported.

Hypothesis 2 stated that the empirically keyed scales would provide incremental variance above and beyond that of the personality scales. To test this hypothesis, the six scales of the PEPI were entered in a stepwise regression followed by the six empirical scales in a stepwise block for the hold-out sample only. When the personality scales were entered in the first block, the \( R^2 \) of .04 for the Intellectance scale was significant \( [F(1, 155) = 6.45, p = .01] \). The empirical Sociability scale was the only empirical scale that provided incremental variance, and entering this scale improved the \( R^2 \) to .07 \( [F(2, 154) = 5.39, p = .006] \).
Because the scales for Conscientiousness, Agreeableness, or Emotional Stability were not significant the hypothesis was not supported. Nonetheless, the finding that the empirical Sociability scale provided incremental variance over the personality scales is consistent with Kluger et al.'s (1991) conclusion that empirical keys measure latent constructs not captured through traditional scoring. As previously noted, many of the items scored for the personality measure were not scored in the biodata scored measure thereby changing the latent structure of the empirical scales.

As a supplementary analysis, the set of twelve scales were entered as a block for the hold-out group to assess the incremental validity of the scored scales. The beta weights and the associated significance tests for the twelve scales are presented in Table 6. The $R^2$ of .11 was not significant ($F(12, 235) = .66$.

Table 6. Regression Results for Absenteeism Variable.

<table>
<thead>
<tr>
<th>Scale (n = 157)</th>
<th>$\beta$ Weight</th>
<th>Significance T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>-.14</td>
<td>.16</td>
</tr>
<tr>
<td>Ambition</td>
<td>-.01</td>
<td>.95</td>
</tr>
<tr>
<td>Likeability</td>
<td>.03</td>
<td>.72</td>
</tr>
<tr>
<td>Intellectance</td>
<td>-.14</td>
<td>.17</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.091</td>
<td>.33</td>
</tr>
<tr>
<td>Sociability</td>
<td>.17</td>
<td>.08</td>
</tr>
<tr>
<td>Neuroticism Bio</td>
<td>-.06</td>
<td>.46</td>
</tr>
<tr>
<td>Ambition Bio</td>
<td>-.001</td>
<td>.98</td>
</tr>
<tr>
<td>Likeability Bio</td>
<td>.02</td>
<td>.85</td>
</tr>
<tr>
<td>Intellectance Bio</td>
<td>.05</td>
<td>.62</td>
</tr>
<tr>
<td>Prudence Bio</td>
<td>-.03</td>
<td>.76</td>
</tr>
<tr>
<td>Sociability Bio</td>
<td>-.13</td>
<td>.12</td>
</tr>
</tbody>
</table>

$p = .79$ and none of the Beta weights were significant ($p > .08$). As such, the results of the supplementary analysis do not support the hypothesis.

Hypothesis 3 employed PAMS methodology to compare the personality profiles of the low absenteeism group to the profile of the high absenteeism group. According to Davison and colleagues
(Davison et al., 1996, Davison & Skay, 1991), PAMS is an exploratory procedure designed to reveal the structure of the prototypical profile of a given sample as well as the degree of correspondence of individuals in the sample to that profile. In essence, the emphasis on PAMS is test or scale variation rather than person variation (Davison & Skay, 1991). To assess scale variation, the PAMS methodology generates a small number of dimensions to produce prototypical profiles. These dimensions are standardized to have a mean of zero and the prototypical profiles are represented as deviations about the prototypical profile's mean of zero. Negative dimension scale values indicate scores below the prototypical profile's own mean and positive dimension scale values signify scores above the profile's own mean. PAMS dimension weights are interpreted similarly to the factor loadings of factor analysis because there are no specific or conventional criteria for interpretation (M.L. Davison, personal communication, April 30, 1996). As such, there is a degree of subjectivity regarding the interpretation of weights. Fit of the model is assessed via the STRESS statistic (Kruskal, 1964). STRESS is a normalized sum of squared errors. Much like factor analysis, PAMS undergoes a number of transformations in order to reduce squared errors through optimal rescaling and iterations of the data computed from the scale value estimates. STRESS equals 0.0 if the model can account for the data perfectly and increases toward 1.0 as the model fits the data less well.

To aid in interpretation and facilitate comparisons, the personality and scored dimensions for the low and high absenteeism groups are combined in Table 7 and profiles are depicted in Figures 1a through 1d on page 67. Each profile will be discussed separately below. For the PEPI, a two-dimensional solution was determined to provide a good fit of the data (STRESS=.06) for the low absenteeism group. Figure 1a depicts one profile for individuals who exhibit low rates of absenteeism to be characterized by positive scores for Prudence (Conscientiousness) and Intellectance (Openness to Experience), and negative scores for Ambition (Extraversion) and Likeability (Agreeableness). In other words, the prototypical low absenteeism individual may be described as being conscientious and enjoying cultural and intellectual pursuits, but is not extraverted nor easygoing with others. A second
low absenteeism profile contained positive elevations for Intellectance and Adjustment (Emotional Stability), and negative scores for Prudence. These results suggest that a second low absenteeism prototypical profile is one in which individuals are open to experience and are well adjusted but do not have a strong work ethic. Partial support for the hypothesis that Conscientiousness, Openness to Experience, and Emotional Stability will be related to absenteeism can be inferred from the elevations for Prudence, Adjustment, and Intellectance. Although not hypothesized, the influence of Ambition (Extraversion) on the first profile may provide indirect evidence for the hypothesis because Hogan has interpreted this factor as being most related to the Conscientiousness dimension.

Figure 1b shows the dimension scores and Table 7 presents the weights for the high absenteeism group. A two-dimensional solution was appropriate for this group as indicated by the STRESS statistic.
Figures 1a - 1d. Sample One Absenteeism PAMS Profiles for Personality and Biodata.
of .01. For the high absenteeism group, one dimension was characterized by negative scores for Adjustment (Emotional Stability), with positive scores for Prudence (Conscientiousness) and Sociability (Extraversion). In other words, a high absenteeism profile could be described by individuals who were likely to be less well adjusted, be outgoing, and have a high work ethic. A second dimension had a positive elevation for Likeability (Agreeableness) and a negative elevation for Intellectance (Openness to Experience), suggesting that people who are easy-going, and enjoy intellectual and cultural activities may describe a profile of the driver who is frequently absent.

The following differences are noted when comparing the high and low absenteeism groups. For the high absenteeism group, the profiles are driven primarily by a combination of either Adjustment, Prudence, and Sociability, or a combination of Likeability, and Intellectance. In contrast, the profiles of the low absenteeism group are characterized by either a combination of Prudence, Intellectance, Likeability, and Ambition, or Prudence, Intellectance, and Adjustment. These overall differences between the personality profiles of the low and high absenteeism groups provide partial support for Hypothesis 3.

Hypothesis 4 predicted that the profiles developed for the empirical scales would have more pronounced elevations for Conscientiousness, Emotional Stability, and Openness to Experience than for traditional scales. The PAMS developed for the low absenteeism group produced the dimensions presented in Table 7 and Figure 1c. As indicated by a STRESS value of .003, a two-dimensional model was appropriate for the data. Here, one dimension had a positive score for the Adjustment (Emotional Stability) scale and negative scores for the Ambition (Extraversion) and Prudence (Conscientiousness) scales, whereas a second dimension had positive scales for Sociability (Extraversion) and negative scores for Ambition. As such, a low absenteeism individual may be well adjusted, less ambitious, and have a low work ethic; or be extraverted with little ambition. In either case, the profiles did not contain elevations for Emotional Stability and Openness to Experience and hence the hypothesis was not supported.
For the high absenteeism group, a two-dimensional model for the biodata scored scales did not provide as adequate a fit to the data (STRESS = .13), as it did for the low absenteeism group. Table 7 and Figure 1d show the dimension scores and profiles for biodata scoring of the PEPI scales. The first dimension yielded positive scores for Likeability (Agreeableness) and Intelectance (Openness to Experience), and negative scores for Sociability (Extraversion) and Adjustment (Emotional Stability). The second dimension had a single elevation for Ambition (Extraversion). Here, the prototypical profiles can be described as likeable, open to experience, less extraverted, and less well adjusted; or ambitious and status seeking. In either case, the results do not support the prediction that the PAMS elevations for Conscientiousness, Agreeableness, and Openness to Experience would be more pronounced for biodata scoring. However, these results demonstrate that the personality profiles differed for low and high absenteeism groups using biodata rather than traditional scoring. These findings further contribute to Kluger et al.'s (1991) conclusion that biodata scored scales measure latent variables different from those measured by traditional personality scales.

Hypothesis 5 proposed that the level parameter for the biodata scored personality measure would be a better predictor of absenteeism than the personality scored measure. Davison and Skay (1991) stated that the level parameter provides an "overall" assessment of the construct or aptitude being measured. The level parameter is an ipsatized standard score taken across all measures. Davison (1994) has referred to the level parameter in personality as a general distress/complaint factor, intensity, or acquiescence. To test Hypothesis 5, one level parameter was developed for personality and a second was developed for biodata. These parameters were then entered into a regression equation. The level parameter for personality was entered in the first step and the resulting $R^2$ of .002 was not significant $[F(1, 469) = .983]$. However, when the level parameter was entered for the empirical scales, the resulting $R^2$ of .09 was significant $[F(2, 469) = 24.39, p = .001]$ supporting the hypothesis.

According to Davison and colleagues (Davison et al., 1996; Davison & Skay, 1991), the PAMS correspondence weights are a measure of association between a given individual and the prototypical
profile, and provide a method of studying the relationship between profiles and criterion variables.

Subsequently, Hypothesis 6 stated that the members of the low absenteeism group would match the predominant profiles better than would the high absenteeism group. Davison et al. (1996) contended that an appropriate test in this case would be to statistically compare the mean person correspondence weights for one group to the weights of the second comparison group to test that a given profile is equally prominent in the two groups. For the personality measures the mean correspondence weights for the first profile dimension for the low and high absenteeism group were -.01 (s.d. = .39) and .04 (s.d. = .38), respectively. These means were not significantly different \( F(1,327) = 1.53, p = .22 \).

Similarly, the mean correspondence weights for the second profile for the low \( (X = -.03, \text{s.d.} = .47) \) and high \( (X = .05, \text{s.d.} = .52) \) absenteeism group did not differ significantly \( F(1,327) = 2.33, p = .13 \).

For empirical scales, mean correspondence weights did not differ significantly for the two groups. The low absenteeism group mean of .03 (s.d. = .39) and the high absenteeism group mean of -.003 (s.d. = .38) did not significantly differ \( F(1,327) = .54, p = .46 \) for profile one; and the mean correspondence weights for the low \( (X = .01, \text{s.d.} = .41) \) and high \( (X = .03, \text{s.d.} = .42) \) absenteeism group did not significantly differ \( F(1,327) = .41, p = .52 \) for profile two.

Hypothesis 7 proposed that the empirical Conscientiousness (Prudence) scale would be a better predictor of absenteeism than empirically derived scales for Emotional Stability or Agreeableness. To conduct this analysis a set of a priori contrast weights was developed according to procedures outlined by Meng et al. (1992). For this analysis, Conscientiousness was assigned a weight of "2," Emotional Stability (Adjustment) and Agreeableness (Likeability) a weight of "-1," and the other scales a weight of zero. The result of this analysis conducted on the hold-out sample \( Z = -.522, p = .72 \) did not support the hypothesis.

Hypothesis 8 proposed that the empirical weights for the Conscientiousness (Prudence) scale would be greater than the other scales. Only items with weights greater than zero that contributed to the scale were compared and the weights were converted to absolute values to control for positive and negative
values. The mean, standard deviation, and minimum and maximum of the weights of each scale are presented in Table 8. An ANOVA comparing the mean weights showed that there were no significant differences among the six scales (F(5, 107) = .31, p = .90).


<table>
<thead>
<tr>
<th>Scale</th>
<th>( \bar{x} )</th>
<th>s.d.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment</td>
<td>1.48</td>
<td>.75</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>Ambition</td>
<td>1.29</td>
<td>.47</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>Likeability</td>
<td>1.31</td>
<td>.63</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>Intellectance</td>
<td>1.5</td>
<td>.86</td>
<td>-4</td>
<td>4</td>
</tr>
<tr>
<td>Prudence</td>
<td>1.47</td>
<td>.84</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>Sociability</td>
<td>1.53</td>
<td>.74</td>
<td>-3</td>
<td>3</td>
</tr>
</tbody>
</table>

The predictive validity of personality and empirical scales was also assessed using a turnover criterion. Similar to Hypothesis 1 for the absenteeism measure, Hypothesis 9 proposed that a significant relationship would exist between turnover and Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability and these relationships would be greater using empirical rather than traditional scoring. The correlations between the six scales of the PEPI using traditional and biodata scoring and turnover for the development and hold-out samples are presented in Table 9. All tests of significance were conducted at a conservative level (p = .008) to control for the number of correlations (.05/6) being tested (Cohen & Cohen, 1983).

For the personality measures, none of the scales were significantly associated with turnover using the .008 significance level, and only Adjustment (Emotional Stability) reached significance at the conventional level for the hold-out sample (r = .15, p = .03). For biodata scoring, all of the scales were significant for the development group (p < .001); but none were significant for the cross-validation sample. However, Likeability (Agreeableness, r = -.15, p = .03) was significant at the conventional level. A total score was also developed for the turnover criterion by summing the weighted item.
Table 9. Correlations between Personality and Biodata PEPI Scales with Turnover Criterion.

<table>
<thead>
<tr>
<th></th>
<th>Personality</th>
<th>Biodata Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hold-Out (n = 157)</td>
<td>Development (n = 315)</td>
</tr>
<tr>
<td>Adjustment</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>Ambition</td>
<td>.02</td>
<td>-.03</td>
</tr>
<tr>
<td>Likeability</td>
<td>.04</td>
<td>-.01</td>
</tr>
<tr>
<td>Intellectance</td>
<td>-.01</td>
<td>-.04</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.07</td>
<td>.03</td>
</tr>
<tr>
<td>Sociability</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td>Total</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

** p < .008, * p < .01

responses and this empirically derived scale predicted turnover in the hold-out sample (r = -.17), albeit not significantly (p = .017). In short, Hypothesis 9 was not supported.

The six pairs of correlations for traditional and biodata scoring were tested for significant differences. The number of significance tests being conducted was also controlled in this analysis and the experimentwise error rate was set at .008 (.05/6). Results from this analysis indicated that the pairs of correlations for Adjustment (Emotional Stability; Z = -2.02, p = .02), Likeability (Agreeableness; Z = -1.79, p = .04), and Sociability (Extraversion; Z = -1.73, p = .04) were not significantly different at the prescribed significance level. Had the less conservative statistical significance level of .05 been used, the results for Likeability would have supported the hypothesis, but the results for Emotional Stability would have been contrary to prediction as the traditionally scored measure was a better predictor of turnover than empirical scoring for the cross-validation group. Finally, the Sociability scales were not hypothesized to differ, but here the argument could be made that the results support the hypothesis because this personality dimension is closely associated with the Likeability (Agreeableness) dimension (Hogan & Hogan, 1992).
Hypothesis 10 employed the PAMS methodology discussed earlier to compare the prototypical profiles derived from the personality scales of those who left the organization to those who remained. Table 10 and Figure 2a on page 74 provide the dimension scores and scale configurations for those who remained ("stayers"). A STRESS statistic of .008 indicated that a two dimensional configuration fit the data well. For those who remained in the organization, one profile contained negative elevations for Ambition (Extraversion) and Likeability (Agreeableness), and a positive elevation for Intellectance (Openness to Experience). In other words, the prototypical profile of a driver who stayed with the organization could be described as one who is less ambitious, low in likeability, but enjoys intellectual and cultural pursuits. A second profile contained a positive elevation for Adjustment (Emotional Stability) combined with negative elevations for Prudence and Sociability. The prototypical personality

<table>
<thead>
<tr>
<th>Scale</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stayers (n = 329)</td>
<td>Quitters (n = 143)</td>
<td>Stayers</td>
<td>Quitters</td>
</tr>
<tr>
<td>Adjustment</td>
<td>.38</td>
<td>.07</td>
<td>1.30*</td>
<td>-.48</td>
</tr>
<tr>
<td>Ambition</td>
<td>-1.09*</td>
<td>.67</td>
<td>.43</td>
<td>.13</td>
</tr>
<tr>
<td>Likeability</td>
<td>-1.10*</td>
<td>1.47*</td>
<td>.51</td>
<td>-.64</td>
</tr>
<tr>
<td>Intellectance</td>
<td>1.69*</td>
<td>-1.58*</td>
<td>.18</td>
<td>-1.13*</td>
</tr>
<tr>
<td>Prudence</td>
<td>.89</td>
<td>-1.44*</td>
<td>-1.33*</td>
<td>1.18*</td>
</tr>
<tr>
<td>Sociability</td>
<td>-.78</td>
<td>.82</td>
<td>-1.11*</td>
<td>.93</td>
</tr>
<tr>
<td>Bio-Adjustment</td>
<td>1.20*</td>
<td>1.17*</td>
<td>.85</td>
<td>.22</td>
</tr>
<tr>
<td>Bio-Ambition</td>
<td>-1.52*</td>
<td>-1.55*</td>
<td>.07</td>
<td>.16</td>
</tr>
<tr>
<td>Bio-Likeability</td>
<td>-1.16*</td>
<td>-1.31*</td>
<td>-.97</td>
<td>-.45</td>
</tr>
<tr>
<td>Bio-Intellectance</td>
<td>1.82*</td>
<td>.52</td>
<td>-.89</td>
<td>1.71*</td>
</tr>
<tr>
<td>Bio-Prudence</td>
<td>.31</td>
<td>1.31*</td>
<td>.15</td>
<td>-.78</td>
</tr>
<tr>
<td>Bio-Sociability</td>
<td>-.65</td>
<td>-.14</td>
<td>.78</td>
<td>-.84</td>
</tr>
</tbody>
</table>

Note: *Elevation is greater than one and is considered in the profile interpretation.
Figures 2a - 2d. Sample One PAMS Profiles for Personality and Biodata Dimension Scores for Turnover.
profile of those who stay in the organization can be described as being well adjusted, low in conscientiousness, and low in extraversion.

A two-dimensional profile also provided a good fit to the data (STRESS = .03) for those who quit the organization ("quitters"). These dimensions shown in Table 10 and Figure 2b depict one dimension that included a positive elevation for Likeability (Agreeableness), and negative elevations for Intellectance (Openness to Experience) and Prudence, and a second dimension that included a negative elevation for Intellectance and a positive elevation for Prudence. As such the prototypical profile of someone who left the organization is (a) someone who is likeable, has little interest in intellectual and cultural pursuits, and has a low work ethic or (b) someone who has a high work ethic but lacks interest in intellectual pursuits.

Taken together, the findings provide indirect support for the hypothesis. The derived profiles differed for those who remained in an organization compared to those who left. Most notably, the profiles diverged in dimension one such that stayers were low in Ambition, high in Intellectance, and low in Likeability, whereas quitters were low in Conscientiousness, low in Intellectance, and high in Likeability. Similarly for dimension two the profile for stayers was driven by high Adjustment, low Prudence, and low Sociability, whereas the quitters’ profile was driven primarily by high Prudence and low Intellectance.

Hypothesis 11 stipulated that the profiles derived from biodata scoring would contain more pronounced elevations for Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability than traditional scoring. Table 10 and Figure 2c present the dimensions for the biodata scored scales for those that remained in the organization. As indicated by the STRESS statistic of .004, a two-dimensional model was appropriate for the data. One dimension was driven primarily by negative elevations for Ambition (Extraversion) and Likeability (Agreeableness), and positive elevations for Adjustment and Intellectance (Openness to Experience). A second dimension contained elevations that differed minimally from the mean of zero with the exception of a positive elevation for Intellectance.
In terms of the hypothesis, only the elevation for Adjustment (Emotional Stability) was more pronounced using biodata scoring than personality scoring for Dimension two. However, for dimension one the elevations for the personality scales were higher than the biodata counterparts. As such, the findings do not support the hypothesis. Nonetheless, these results contribute further to the notion that biodata scoring of personality items measure latent constructs different than those measured with traditional scoring (Mumford & Owens, 1987; Mumford & Stokes, 1992).

For those who did not remain in the organization a two-dimensional configuration did not meet the criteria of .08 (Davison & Skay, 1991) to provide a good fit to the data (STRESS = .10). Because the profiles derived for this group did conform to the two-dimensional configuration with personality scoring but not biodata scoring, offers additional support that biodata scoring changes the latent structure of personality tests (Kluger et al., 1991). To remain consistent with the previous profiles, a two-dimensional interpretation was retained to aide in comparing the profiles for personality and biodata scoring. The profiles and dimension scores are presented in Figure 2d and Table 10, respectively. For individuals who left the organization, one profile contained positive elevations for Adjustment (Emotional Stability) and Prudence (Conscientiousness), and negative elevations for Ambition (Extraversion) and Likeability (Agreeableness). A second profile dimension was driven primarily by a positive elevation for Intellectance.

The comparison of stayers versus quitters shows differences in dimension one for Prudence and Intellectance. The quitters' profile had an elevation for the Prudence biodata scored scale that the stayers' profile did not, and the stayers' profile had an elevation for Intellectance that the quitters' profile did not. The most dramatic difference between stayers and quitters for dimension two can be seen in the Intellectance biodata scored scale. For stayers, the Intellectance scale scores were just below the mean, whereas for quitters the Intellectance scale had a positive elevation.

Hypothesis 12 posited that the level parameter for the biodata scored personality measure would explain more variance in turnover than the level parameter for traditional scoring. For this analysis the
turnover variable was treated as a continuous rather than dichotomous variable allowing the use of linear regression rather than logistic regression. Rationale for this approach can be found in Barrick and Mount’s (1996) analysis in which the Big Five dimensions were regressed onto a turnover variable that the authors treated as continuous. Additional support comes from Campion (1991) who stated that turnover measures may best be described as continuous theoretical constructs because decisions to leave are a function of both the organization and the employee. Campion argued that an employee voluntarily deciding to leave may be placed on a continuum ranging from completely voluntary turnover where the employee quits because of better pay somewhere else, to mutual agreement where the employee agrees to quit because of a disagreement with management, to involuntary turnover due to dismissal. To be consistent with Barrick and Mount’s analysis, linear regression was deemed the appropriate statistical test for this study and the level parameters for the personality measure where regressed onto the turnover measure followed by the biodata scored level parameters. Results from the regression did not support the hypothesis. The level parameter for the personality score was not significant and the biodata scored level parameter was only marginally significant \[F(1, 155) = 3.21, p = .08\].

Summary

Overall, Sample One results do not provide strong support for the hypotheses. First, the specified personality factors of Conscientiousness, Agreeableness, and Emotional Stability did not prove to be reliable predictors of absenteeism. Second, the correlations for Emotional Stability and Openness to Experience were not consistently significant for the hold-out and developmental samples. Third, empirical scales that were significant for the developmental samples were not significant for the hold-out sample as only the Sociability biodata scored scale and the Total biodata scale approached significance for the hold-out sample. Fourth, the hypotheses touting the superiority of the Conscientiousness construct were not supported in this sample. Finally, the hypotheses that the biodata
scored scales would provide incremental variance above and beyond the personality scales was also not supported.

The personality profiles for the high and low absenteeism groups did differ, but the low absenteeism group did not contain the elevations specified by the hypothesis. For the low absenteeism group, dimension one contained positive elevations for Openness to Experience and Conscientiousness and negative elevations for Agreeableness and Extraversion. Dimension two contained a positive elevation for Neuroticism coupled with a negative elevation for Openness to Experience and a negative elevation for Conscientiousness.

The biodata scored profiles did not contain more pronounced elevations for the Neuroticism, Agreeableness, and Conscientiousness scales as predicted. Nonetheless, the biodata scored profiles differed from the personality profiles in the content and direction of the elevations. Dimension one for the low absenteeism group had a positive elevation for Neuroticism, a negative elevation for Extraversion, and a negative elevation for Conscientiousness. Dimension two had a negative elevation for one component of Extraversion for the PEPI (Ambition) and a positive elevation for the second component of Extraversion (Sociability).

As indicated above, the form of neither the empirical nor personality profiles conformed to those specified by the hypotheses. A second analysis of the profiles was conducted by examining the overall level of personality and biodata scored scales as measured by the level parameters. This hypothesis was supported such that the overall component of biodata scoring was a better predictor of absenteeism than the level parameter for personality. A third analysis comparing the profiles' mean correspondence weights for the low and high absenteeism groups did not produce significant differences.

When turnover was used as the criterion variable, none of the personality scales were predictive. For the biodata scored scales, only the scales for Agreeableness and the Total scale were marginally significant. In combination, the results do not demonstrate the importance of Conscientiousness in the prediction of turnover.
The personality profiles for the stayer contained a positive elevation for Openness to Experience as predicted, but had negative elevations for Agreeableness and Extraversion that were not. A second personality dimension was also contrary to prediction with negative elevations for Conscientiousness and Extraversion and a positive elevation for Neuroticism.

The profiles developed for the biodata scored scales did not contain more pronounced elevations as specified. Instead, one dimension of the stayer profile contained positive elevations for Emotional Stability and Openness to Experience, but negative elevations for Extraversion and Likeability. The second dimension of the stayer profile contained elevations that did not deviate from the mean and hence could not be interpreted.

In sum, the results for this sample did not contribute to the literature demonstrating the utility of Conscientiousness construct. However, the results do show that biodata scoring can make a unique contribution to the prediction of absenteeism behavior. Although the exact forms of the personality and biodata scored profiles were not confirmed, the overall assessment of biodata scoring as measured by the PAMS level parameter did predict absenteeism behavior.

SAMPLE TWO

The second analysis used an aggregate sample of bus drivers from transit agencies located throughout the United States. Because the data collected for this sample did not include turnover data only Hypotheses 1 through 8 were analyzed. To develop empirical scales, the continuous absenteeism criterion variable was split into three approximately equivalent groups. Individuals with 8 or fewer days of absence during the year comprised the low absenteeism group (n = 250) and individuals with 31 or more days of absence during the year were included in the high absenteeism group (n = 258). According to Conti et al. (1996) driver absenteeism results in significant costs to the organization and most transit agencies have taken measures to reduce the occurrence of absences.

The original sample was split into hold-out and development groups to derive empirical scales. These groups were compared for differences in demographic variables, differences in mean scale scores
of the six HPI scales, and differences in the number of days absent to mitigate the influence these variables may have on empirical item weights. The results from the chi-square analyses of demographic differences presented in Table 11 show that there were no significant differences between groups.

Table 11. HPI Sample Characteristics as a Percentage of Total.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>White</th>
<th>Hispanic</th>
<th>Black</th>
<th>Asian</th>
<th>Other</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross</td>
<td>76.2</td>
<td>23.8</td>
<td>33.3</td>
<td>12.3</td>
<td>46.0</td>
<td>3.6</td>
<td>3.2</td>
<td>1.6</td>
</tr>
<tr>
<td>Develop</td>
<td>78.6</td>
<td>21.4</td>
<td>36.1</td>
<td>11.4</td>
<td>48.0</td>
<td>2.0</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>(\chi^2) Value</td>
<td>(1, 0.55) = .46</td>
<td>(5, 5.64) = .34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similarly, group means did not differ for the scales of the HPI or in the mean number of absences for the year 1995. These results are shown in Table 12.

Table 12. Group Comparison for Absenteeism or HPI Scale Differences.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Hold-Out (n=248)</th>
<th>Development (n=480)</th>
<th>F-Test (1,726)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences</td>
<td>(\bar{X} = 24.89,) s.d. = 26.27</td>
<td>(\bar{X} = 25.48,) s.d. = 26.55</td>
<td>.03, p = .86</td>
</tr>
<tr>
<td>Likeability</td>
<td>(\bar{X} = 5.76,) s.d. = 3.18</td>
<td>(\bar{X} = 5.86,) s.d. = 2.96</td>
<td>.22, p = .63</td>
</tr>
<tr>
<td>Sociability</td>
<td>(\bar{X} = 16.23,) s.d. = 5.63</td>
<td>(\bar{X} = 15.87,) s.d. = 5.59</td>
<td>.67, p = .41</td>
</tr>
<tr>
<td>Intellectance</td>
<td>(\bar{X} = 16.17,) s.d. = 6.20</td>
<td>(\bar{X} = 16.43,) s.d. = 5.91</td>
<td>.31, p = .40</td>
</tr>
<tr>
<td>Prudence</td>
<td>(\bar{X} = 23.32,) s.d. = 5.29</td>
<td>(\bar{X} = 22.76,) s.d. = 5.38</td>
<td>1.82, p = .18</td>
</tr>
<tr>
<td>Ambition</td>
<td>(\bar{X} = 24.78,) s.d. = 5.98</td>
<td>(\bar{X} = 24.40,) s.d. = 6.43</td>
<td>.60, p = .44</td>
</tr>
<tr>
<td>Adjustment</td>
<td>(\bar{X} = 14.27,) s.d. = 8.55</td>
<td>(\bar{X} = 13.26,) s.d. = 8.28</td>
<td>.37, p = .54</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
All of the HPI items were not used for the development of the empirical scales. Only 68.4% (18 of 26) of the Likeability items, 71.8% (23 of 32) of the Sociability items, 64.5% (20 of 31) of the Intellectance items, 60.6% (20 of 33) of the Prudence items, 52.8% (19 of 36) of the Ambition, and 59.2% (29 of 49) of the Adjustment items were used in the empirical scales. The reliabilities for the scales are consistent with those reported by Mumford and Owens (1987) who stated that biodata scored scales frequently measure multi-dimensional constructs with relatively few items. The reliabilities for Ambition (α = .40), Intellectance (α = .40), Likeability (α = .10), Sociability (α = .10), Prudence (α = .22), and Adjustment (α = .15) were all below those reported for the personality scales.

Hypothesis 1 posited that the relationship between Conscientiousness, Agreeableness, and Emotional Stability and absenteeism would be greater using empirical scoring rather than traditional scoring. The Sample Two correlations between the personality scales and biodata scored scales with absenteeism are presented in Table 13 and all correlations were tested at the .008 level of significance to control for the number of correlations being tested (.05/6). Only the Likeability scale approached significance for the development group (r = .10, p = .02), and none of the other personality scales were significant for either the development sample or hold-out samples. When the personality items were

<table>
<thead>
<tr>
<th>Personality</th>
<th>Biodata Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hold-Out (n=248)</td>
<td>Development (n=480)</td>
</tr>
<tr>
<td>Adjustment</td>
<td>.04</td>
</tr>
<tr>
<td>Ambition</td>
<td>.06</td>
</tr>
<tr>
<td>Likeability</td>
<td>-.01</td>
</tr>
<tr>
<td>Intellectance</td>
<td>.02</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.01</td>
</tr>
<tr>
<td>Sociability</td>
<td>.01</td>
</tr>
<tr>
<td>Total</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Note: ** p < .008, * p < .01
empirically scaled, the scales for Adjustment (Emotional Stability, \( r = -0.15 \)), Likeability (\( r = -0.12 \)), Sociability (Extraversion, \( r = -0.13 \)), and the Total scale (\( r = -0.22 \)) were significant for the development sample (\( p < .004 \)) but none were significant for the hold-out sample. However, scales for Adjustment (\( r = -0.12, p = .03 \)) and Total (\( r = -0.13, p = .02 \)) were significant at the .05 level of significance.

Similar to Sample One, these results do not offer support for the hypothesis.

The relationship between the traditionally scored and empirically scored scales were tested to assess the superiority of the empirical approach over the traditional method using the Fisher Z transformation discussed earlier. Although Conscientiousness (Prudence), Agreeableness (Likeability), and Emotional Stability (Adjustment) were the primary focus of the study, all six pairs of cross validated correlations were tested for significance for the hold-out sample. The results of the Z (normal curve) test and the 95% confidence intervals are presented in Table 14. Here, the correlations did not differ significantly in magnitude. Only the comparison of scale correlations for the Adjustment (Emotional Stability) scale approached significance (\( Z = -1.47, p = .07 \)) at the conventional .05 significance level. Therefore, support for Hypothesis 1 was not demonstrated in this sample.

Hypothesis 2 proposed that the biodata scored scales would provide incremental variance beyond that of the traditional scales. Consistent with the analysis used for Sample One, the set of personality scales were entered stepwise into the equation followed by the empirically scored scales. None of the personality scales predicted absenteeism, and only the Adjustment (Emotional Stability) biodata scored

<table>
<thead>
<tr>
<th>Scale Compared</th>
<th>Z Score</th>
<th>Significance</th>
<th>Upper CI</th>
<th>Lower CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment</td>
<td>-1.47</td>
<td>( p = .07 )</td>
<td>.019</td>
<td>-.333</td>
</tr>
<tr>
<td>Ambition</td>
<td>-.40</td>
<td>( p = .343 )</td>
<td>.343</td>
<td>-.110</td>
</tr>
<tr>
<td>Likeability</td>
<td>-.71</td>
<td>( p = .240 )</td>
<td>.098</td>
<td>-.245</td>
</tr>
<tr>
<td>Intellectance</td>
<td>-.567</td>
<td>( p = .285 )</td>
<td>.073</td>
<td>-.151</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.430</td>
<td>( p = .333 )</td>
<td>.131</td>
<td>-.224</td>
</tr>
<tr>
<td>Sociability</td>
<td>-.90</td>
<td>( p = .184 )</td>
<td>.076</td>
<td>-.260</td>
</tr>
</tbody>
</table>
scale approached significance for the hold-out group \( F(1, 246) = 3.411, p = .07 \). Consequently, this hypothesis was not supported. As a further test, the set of twelve scales were entered as a block for the hold-out group to assess the incremental validity of the biodata scored scales. The beta weights and the associated significance tests for the twelve scales are presented in Table 15. The \( R^2 \) of .04 was not significant \( F(12, 235) = .66, p = .79 \) and none of the Beta weights were significant \( (p > .20) \).

<table>
<thead>
<tr>
<th>Scale (n = 248)</th>
<th>β Weight</th>
<th>Significance T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>.073</td>
<td>.470</td>
</tr>
<tr>
<td>Ambition</td>
<td>.053</td>
<td>.617</td>
</tr>
<tr>
<td>Likeability</td>
<td>-.058</td>
<td>.327</td>
</tr>
<tr>
<td>Intellectance</td>
<td>.025</td>
<td>.800</td>
</tr>
<tr>
<td>Prudence</td>
<td>-.001</td>
<td>.984</td>
</tr>
<tr>
<td>Sociability</td>
<td>-.081</td>
<td>.419</td>
</tr>
<tr>
<td>Neuroticism Bio</td>
<td>-.094</td>
<td>.252</td>
</tr>
<tr>
<td>Ambition Bio</td>
<td>.042</td>
<td>.656</td>
</tr>
<tr>
<td>Likeability Bio</td>
<td>-.057</td>
<td>.441</td>
</tr>
<tr>
<td>Intellectance Bio</td>
<td>.004</td>
<td>.963</td>
</tr>
<tr>
<td>Prudence Bio</td>
<td>-.049</td>
<td>.521</td>
</tr>
<tr>
<td>Sociability Bio</td>
<td>-.060</td>
<td>.419</td>
</tr>
</tbody>
</table>

Hypothesis 3 explored differences between the personality profiles of the high and low absenteeism groups using PAMS methodology. Profiles were generated for the high and low absenteeism groups using traditional and empirical scaling. These dimension weights for the derived profiles for the high and low absenteeism groups are presented in Table 16 and shown in figures 2a through 2d on page 85.

The personality profile for the low absenteeism group conformed to a two-dimension configuration as evidenced by the STRESS value of .03. The profiles for the low absenteeism group using traditional scoring are shown in Figure 2a. The first dimension contained positive elevations for Likeability (Agreeableness) and Adjustment (Emotional Stability), and a negative elevation for Prudence.
Table 16. HPI Personality and Biodata Scored Dimension Scores for High and Low Absenteeism.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Dimension 1</th>
<th></th>
<th>Dimension 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (n = 258)</td>
<td>High (n = 250)</td>
<td>Low (n = 258)</td>
<td>High (n = 250)</td>
</tr>
<tr>
<td>Adjustment</td>
<td>1.89*</td>
<td>1.85*</td>
<td>.14</td>
<td>-.17</td>
</tr>
<tr>
<td>Ambition</td>
<td>-.76</td>
<td>-.64</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Likeability</td>
<td>1.62*</td>
<td>1.48*</td>
<td>-.56</td>
<td>.69</td>
</tr>
<tr>
<td>Intellectance</td>
<td>-.75</td>
<td>-.91</td>
<td>.66</td>
<td>-.83</td>
</tr>
<tr>
<td>Prudence</td>
<td>-1.67*</td>
<td>-1.55*</td>
<td>-.83</td>
<td>1.00*</td>
</tr>
<tr>
<td>Sociability</td>
<td>-.33</td>
<td>-.24</td>
<td>.59</td>
<td>-.69</td>
</tr>
<tr>
<td>Bio-Adjustment</td>
<td>1.20*</td>
<td>1.17*</td>
<td>.85</td>
<td>.22</td>
</tr>
<tr>
<td>Bio-Ambition</td>
<td>-1.52*</td>
<td>-1.55*</td>
<td>.07</td>
<td>.16</td>
</tr>
<tr>
<td>Bio-Likeability</td>
<td>-1.16*</td>
<td>-1.31*</td>
<td>-.97</td>
<td>-.45</td>
</tr>
<tr>
<td>Bio-Intelectance</td>
<td>1.82*</td>
<td>.52</td>
<td>-.89</td>
<td>1.71*</td>
</tr>
<tr>
<td>Bio-Prudence</td>
<td>.31</td>
<td>1.31*</td>
<td>.15</td>
<td>-.78</td>
</tr>
<tr>
<td>Bio-Sociability</td>
<td>.78</td>
<td>1.38*</td>
<td>.53</td>
<td>.38</td>
</tr>
</tbody>
</table>

Note: *Elevation is greater than one and is considered in the profile interpretation.

(Conscientiousness). A prototypical personality profile of a low absenteeism driver could therefore be described as being well adjusted and likeable, but having a low work ethic. The second dimension did not contain marked elevations for any of the six dimensions and hence did not provide a meaningful interpretation.

The profiles for the high absenteeism group are presented in Figure 2b. A two-dimensional configuration was also appropriate for this data set (STRESS = .001). The first dimension for the high absenteeism group contained a similar pattern of elevations for Likeability, Adjustment, and Prudence to that of the low absenteeism profile. The divergence between the two groups occurred primarily in the second dimension for the Prudence (Conscientiousness) scale, such that the high absenteeism prototypical profile contained a positive elevation for work ethic not present in the low absenteeism...
Figures 3a - 3d. Sample Two PAMS Absenteeism Profiles for Personality and Biodata.
prototypical profile. As such the hypothesis that individuals who are conscientious would be less likely to be absent was not supported.

Hypothesis 4 predicted that the PAMS profiles generated through empirical scoring would be characterized by higher elevations for Emotional Stability, Conscientiousness, Agreeableness, and Openness to Experience than profiles developed through traditional scoring. Once again, biodata scoring changed the profile configurations for both the low and high absenteeism groups. The dimension scores for the low absenteeism group biodata scored scale are presented in Table 16 and depicted in Figure 2c. A STRESS statistic of .13 suggested that a two-dimensional model provided a marginal fit to the data, especially when compared to the two-dimensional model for the traditionally scored scale (STRESS = .03). Here, one dimension exhibited a positive elevation for Ambition (Extraversion) and a negative elevation for Intellectance (Openness to Experience), in other words being outgoing but having little interest in cultural or intellectual pursuits. A second dimension exhibited a negative elevation for Ambition and Sociability and a positive elevation for Likeability (Agreeableness). Therefore the prototypical profile of a prototypical low absenteeism driver could be described as likeable, yet lacking in ambition or extraversion.

A different combination of profiles was generated for the high absenteeism group as shown in Table 16 and Figure 2d. The STRESS statistic of .09 shows that a two-dimensional model provided an adequate fit to the data though not as good a fit as for traditional scoring (STRESS = .001). For the high absenteeism group, one dimension contained positive elevations for Adjustment (Emotional Stability) and Likeability (Agreeableness), and negative elevations for Intellectance and Prudence. The second dimension contained negative elevations for Ambition (Extraversion) and Prudence (Conscientiousness). Although the Hypothesis of greater elevations for biodata scoring was not supported, two conclusions can be drawn: (1) PAMS profiles for the low and high absenteeism groups were characterized by slightly different configurations; and (2) biodata scoring of the personality items
changed the interpretation of profiles as seen in a poorer fit of the model perhaps suggesting that different constructs are being measured.

Hypothesis 5 proposed that the level parameter for the biodata scored personality would be a better predictor of absenteeism than the level parameter for the personality measure. As stated earlier, the level parameter is a measure of one's correspondence to the predominant personality profile. Further, because biodata restricts item scoring only to those items that best predict a criterion, one would expect the level parameter for biodata scoring to be most closely associated with the criterion. Only the hold-out sample was used for this analysis. When the level parameter for personality (X = .04, s.d. = .28) was entered into the regression equation it was not significant [F(1, 246) = .33, p = .56]. However, entering the biodata scored level parameter (X = -.11) improved the R² to .01 but it also was not significant [F(2, 245) = 1.51, p = .22]. Hence the hypothesis was not supported.

Hypothesis 6 stated that members of the low absenteeism group would match the predominant profiles better than members of the high absenteeism group. This analysis was performed using the hold-out sample. Here, the mean person correspondence weights for the low absenteeism group were compared to those of the high absenteeism group to assess whether the predominant profile is equally prominent for both groups. For the personality measures the mean correspondence weights for the first profile dimension for the low and high absenteeism group were -.03 (s.d. = -.03) and -.01 (s.d. = .52) respectively. These means were not significantly different [F(1, 171) = .03, p = .86]. Similarly, the mean correspondence weights for the second profile for the low (X = -.001, s.d. = .65) and high (X = -.015, s.d. = .70) absenteeism group did not differ significantly [F(1, 171) = .02, p = .89].

The correspondence weights for the biodata scored scales were also compared for the hold-out sample. For the first profile, the mean correspondence weights for the low (X = .05, s.d. = .35) and the high (X = .04, s.d. = .36) absenteeism groups did not differ [F(1, 171) = .04, p = .84]. The correspondence means for the low (X = .04, s.d. = .65) and high (X = -.06, s.d. = .57) absenteeism groups also did not differ for the second dimension profile [F(1, 171) = 1.14, p = .29].
Hypothesis 7 proposed that the empirical Conscientiousness (Prudence) scale would be a better predictor of absenteeism than empirically derived scales for Emotional Stability or Agreeableness. A priori contrast weights were developed using the same Meng et al. (1992) procedure used with Sample One. Conscientiousness was assigned a weight of “2,” Emotional Stability (Adjustment) and Agreeableness (Likeability) a weight of “-1,” and the remaining scales were assigned a weight of zero. The result of this analysis ($Z = .171, p = .57$) for the hold-out sample did not support the hypothesis.

Hypothesis 8 stated that the empirical weights for the Conscientiousness (Prudence) scale would be greater than the other scales. Table 17 presents the mean, standard deviation, and minimum and maximum of the weights. Prior to the analysis, weights were converted to absolute values to control for positive and negative values. An Anova comparing the mean weights did not support the hypothesis because the weights for the six scales did not significantly differ [$F(5, 160) = 1.25, p = .29$].

**Summary**

The results obtained for Sample Two also did not demonstrate the effectiveness of the Conscientiousness and Agreeableness personality dimensions to predict absenteeism. Only the Adjustment (Emotional Stability) personality scale, and the biodata scored scales for Adjustment and Total predicted absenteeism for the cross-validation sample, albeit not statistically significant because of the stringent level used in this study. Neither the correlation coefficients nor the item weights for Conscientiousness were greater than those for Agreeableness or Emotional Stability. Moreover, the

<table>
<thead>
<tr>
<th>Scale</th>
<th>X</th>
<th>s.d.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjustment</td>
<td>.64</td>
<td>.67</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>Ambition</td>
<td>.58</td>
<td>.58</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>Likeability</td>
<td>1.05</td>
<td>.97</td>
<td>-3</td>
<td>3</td>
</tr>
<tr>
<td>Intellectance</td>
<td>.72</td>
<td>.74</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>Prudence</td>
<td>.75</td>
<td>.75</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>Sociability</td>
<td>.84</td>
<td>.58</td>
<td>-2</td>
<td>2</td>
</tr>
</tbody>
</table>
effectiveness of biodata scoring was not demonstrated because the biodata scored scales did not provide incremental variance beyond that explained by the personality scales.

The dimension one profile for the low absenteeism group did contain positive elevations for Emotional Stability and Agreeableness as specified. However, the negative elevation for Conscientiousness was contrary to prediction. The dimension two profile did not contain a meaningful pattern of elevations for interpretative purposes and therefore did not support the hypothesis. In addition, whereas the biodata scored profiles differed from the personality profiles, the elevations were not more pronounced as predicted. Dimension one contained positive Emotional Stability and Openness to Experience elevations, with negative elevations for Extraversion (Ambition) and Conscientiousness (Prudence). The overall assessment of biodata scoring did not significantly predict absenteeism and it did not provide incremental variance beyond that explained by personality. Similarly, the mean correspondence weights for the low and high absenteeism groups did not differ suggesting that both conformed equally well to the profiles.

In sum, the results suggest that biodata scoring not only changes the pattern of scoring as evidenced by divergent profiles, but that it can sometimes improve prediction. This is true despite the low internal consistencies of the scales as shown by the low Cronbach's alphas of the empirical scales. The biodata scored scales for Sample Two were developed using a sample of drivers from across agencies and the data were collected in a concurrent rather than predictive manner that may have reduced the ability of the biodata scored scales to cross-validate (e.g., Stokes et al., 1993). For example, many of the items did not discriminate well between the high and low absenteeism group in the developmental sample and hence were not included in the scale. As such, a larger pool of items may result in higher correlations between predictor and criterion based on chance (Mumford & Owens, 1987). Sample Three attempted to overcome some of Sample Two's limitations by using applicant data instead of incumbent data, collecting data from one specific agency, and employing a larger set of
items. An additional advantage of Sample Three is that occupations other than bus drivers (transit or school) were used to develop and cross-validate scales.

SAMPLE THREE

Scale Development

Empirical scales were developed for each year that absenteeism data were collected, beginning January 1, 1991 and ending December, 1996. Each successive year of the sample included absenteeism data for new employees hired by the organization during that year and incumbents who had completed the CPI in previous years but excluded employees who had left the organization. For example, an employee may have been hired by the organization in 1992 and remained in the organization until 1995. This employee would therefore have absenteeism data for 1992 through 1995.

For each year, absenteeism scales were developed using two-thirds of the sample as the development group and the remaining third of the sample as the cross validation group. Scales were developed using the contrasting groups approach discussed earlier. The high absenteeism group for all six years consisted of individuals who used more than 93.5 hours of absenteeism per year. The low absenteeism group was comprised of employees who used less than 38.3 hours per year. These number were obtained by averaging the number of hours used for each of the six years for a one-third high and one-third low absenteeism group split.

Cronbach’s alphas were computed for each empirical scale developed for the six years of absenteeism data. Table 18 on the following page shows the \( \alpha \) for each scale and the percentage of the original items used to develop that scale. As shown, the reliabilities of the personality scales were relatively consistent from year to year. The internal consistencies of the biodata scored scales, on the other hand, ranged from a low of .40 to a high of .74. These findings are similar to those reported by Mumford and Owens (1987) and are consistent with the line of reasoning that biodata scored scales are multidimensional. In addition, the biodata scored scales did not contain the exact number of items used for the personality scales. The number of items used in the scoring of the empirical scales varied.
from a low of 40% of the Neuroticism items for the 1996 scale to a high of 64.9% of the items for the 1992 Conscientiousness scale.

Because of these differences, the decision was made to construct biodata scored scales for each year rather than to aggregate the data to develop a scale that would predict an average yearly absenteeism figure. The primary reason was congruent with the focus of the study to assess the effectiveness of the technique of applying biodata scoring to personality measures. Using this methodology, six samples containing a different composition and number of participants were used to develop and cross validate the scales. Analyzing the data in this manner allowed for a means of internal replication because biodata scoring should improve validity irrespective of the year that the scales were developed. The six samples were tested independently for differences in demographic variables, mean scale score differences for the five scales of the CPI and the derived biodata scored scales, and mean differences for the absenteeism measure.

<table>
<thead>
<tr>
<th>Scale</th>
<th>1991</th>
<th>1992</th>
<th>1993</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-α</td>
<td>B-α</td>
<td>%</td>
</tr>
<tr>
<td>New</td>
<td>.87</td>
<td>.44</td>
<td>62.4</td>
</tr>
<tr>
<td>Extra</td>
<td>.81</td>
<td>.33</td>
<td>60.5</td>
</tr>
<tr>
<td>Opn/Exp</td>
<td>.65</td>
<td>.45</td>
<td>53.6</td>
</tr>
<tr>
<td>Agree</td>
<td>.73</td>
<td>.44</td>
<td>57.6</td>
</tr>
<tr>
<td>Consc</td>
<td>.82</td>
<td>.57</td>
<td>64.0</td>
</tr>
</tbody>
</table>

Note: P-α = Personality α; B-α = Biodata α.

<table>
<thead>
<tr>
<th>Scale</th>
<th>1994</th>
<th>1995</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-α</td>
<td>B-α</td>
<td>%</td>
</tr>
<tr>
<td>New</td>
<td>.87</td>
<td>.74</td>
<td>45.6</td>
</tr>
<tr>
<td>Extra</td>
<td>.81</td>
<td>.64</td>
<td>61.8</td>
</tr>
<tr>
<td>Opn/Exp</td>
<td>.70</td>
<td>.47</td>
<td>60.8</td>
</tr>
<tr>
<td>Agree</td>
<td>.73</td>
<td>.55</td>
<td>48.9</td>
</tr>
<tr>
<td>Consc</td>
<td>.81</td>
<td>.63</td>
<td>51.4</td>
</tr>
</tbody>
</table>

Note: P-α = Personality α; B-α = Biodata α.
The chi-square analyses conducted for the six samples presented in Table 19 show that there were no significant differences (p > .05) for the demographic variables. The results for the mean scale differences for the personality scales are presented in Table 20 and the mean scale differences for the biodata scored scales are presented in Table 21. Because of the number of univariate tests being conducted, a probability level of .01 was established to examine differences. Given this criteria, none of the differences were significant. However, had the traditional .05 level of significance been used, the hold-out and development groups would have had 3 significant differences out of 60 predictor variables by chance, and one significant difference out of six criterion comparisons. These differences are as follows: (1) the Conscientiousness mean for the cross validation ($X = 142.1$, s.d. = 8.9) group was

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dev</td>
<td>Cross</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Am/Ind</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filipino</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dev</td>
<td>Cross</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Am/Ind</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filipino</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 20. CPI Personality Mean Scale and Absenteeism Comparisons for Hold-Out and Development Groups.

<table>
<thead>
<tr>
<th>Scale</th>
<th>1991 (sample sizes in parentheses)</th>
<th>1992 (sample sizes in parentheses)</th>
<th>1993 (sample sizes in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dev (362)</td>
<td>Cross (175)</td>
<td>F-Test</td>
</tr>
<tr>
<td>Neurotic</td>
<td>X=130.5</td>
<td>s.d. = 9.2</td>
<td>p = .85</td>
</tr>
<tr>
<td>Extra</td>
<td>X=104.3</td>
<td>s.d. = 7.5</td>
<td>p = .06</td>
</tr>
<tr>
<td>Neurotic</td>
<td>X=130.5</td>
<td>s.d. = 9.2</td>
<td>p = .85</td>
</tr>
<tr>
<td>Extra</td>
<td>X=104.3</td>
<td>s.d. = 7.5</td>
<td>p = .06</td>
</tr>
<tr>
<td>Neurotic</td>
<td>X=130.5</td>
<td>s.d. = 9.2</td>
<td>p = .85</td>
</tr>
<tr>
<td>Extra</td>
<td>X=104.3</td>
<td>s.d. = 7.5</td>
<td>p = .06</td>
</tr>
<tr>
<td>Neurotic</td>
<td>X=130.5</td>
<td>s.d. = 9.2</td>
<td>p = .85</td>
</tr>
<tr>
<td>Extra</td>
<td>X=104.3</td>
<td>s.d. = 7.5</td>
<td>p = .06</td>
</tr>
</tbody>
</table>

significantly higher than the mean for the development group \( \bar{X} = 140.4, \text{s.d.} = 9.1; F(1, 535) = 4.06, p = .04 \) for the 1991 sample, (2) the Neuroticism mean for the cross-validation group \( \bar{X} = 229.4, \text{s.d.} = 9.3 \) differed significantly \( F(1, 1397) = 4.48, p = .03 \) from the development group mean \( \bar{X} = 230.4, \text{s.d.} = 8.8 \) for the 1995 sample, (3) the mean for the 1995 cross validation sample for the Openness to Experience scale \( \bar{X} = 137.5, \text{s.d.} = 6.7 \) also differed significantly \( F(1, 1397) = 4.13, p = .04 \) from
the mean for the development group ($X = 136.7, \text{s.d.} = 6.6$), and (4) the mean number of absences for the 1991 cross-validation group ($X = 77.8, \text{s.d.} = 158.1$) differed significantly [F(1, 533) = 4.47, p = .03] from the mean number of absences for the development group ($X = 53.5, \text{s.d.} = 104.9$). Because these differences occurred for the personality scales but not the biodata scored scales has implications for comparing the correlations between the three 1991 and 1995 personality scales noted above and the

Table 21. CPI Biodata Mean Scale Comparisons for Hold-Out and Development Groups.

<table>
<thead>
<tr>
<th>Scale</th>
<th>1991 (sample sizes in parentheses)</th>
<th>1992 (sample sizes in parentheses)</th>
<th>1993 (sample sizes in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dev (793)</td>
<td>Cross (375)</td>
<td>F-Test</td>
</tr>
<tr>
<td>Bio-Near</td>
<td>$X = 9.1$</td>
<td>$X = 9.5$</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>s.d. = 11.9</td>
<td>s.d. = 11.6</td>
<td>p = .76</td>
</tr>
<tr>
<td>Bio-Extra</td>
<td>$X = 4.9$</td>
<td>$X = 4.0$</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>s.d. = 3.5</td>
<td>s.d. = 3.3</td>
<td>p = .99</td>
</tr>
<tr>
<td>Bio-Open</td>
<td>$X = 65$</td>
<td>$X = 1.34$</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>s.d. = 11.3</td>
<td>s.d. = 11.0</td>
<td>p = .50</td>
</tr>
<tr>
<td>Bio-Agree</td>
<td>$X = 7.2$</td>
<td>$X = 7.8$</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>s.d. = 3.0</td>
<td>s.d. = 11.2</td>
<td>p = .61</td>
</tr>
<tr>
<td>Bio-Cons</td>
<td>$X = 6.4$</td>
<td>$X = 5.5$</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>s.d. = 15.4</td>
<td>s.d. = 16.8</td>
<td>p = .47</td>
</tr>
<tr>
<td>Bio-Total</td>
<td>$X = 4.7$</td>
<td>$X = 7.6$</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>s.d. = 3.9</td>
<td>s.d. = 39.9</td>
<td>p = .42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1994 (sample sizes in parentheses)</th>
<th>1995 (sample sizes in parentheses)</th>
<th>1996 (sample sizes in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>Dev (902)</td>
<td>Cross (294)</td>
</tr>
<tr>
<td>Bio-Near</td>
<td>$X = 31.8$</td>
<td>$X = 32.0$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 14.3</td>
<td>s.d. = 13.1</td>
</tr>
<tr>
<td>Bio-Extra</td>
<td>$X = 1.3$</td>
<td>$X = 1.3$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 13.5</td>
<td>s.d. = 14.7</td>
</tr>
<tr>
<td>Bio-Open</td>
<td>$X = 22.3$</td>
<td>$X = 21.7$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 13.6</td>
<td>s.d. = 13.8</td>
</tr>
<tr>
<td>Bio-Agree</td>
<td>$X = 12.13$</td>
<td>$X = 12.1$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 14.3</td>
<td>s.d. = 14.6</td>
</tr>
<tr>
<td>Bio-Cons</td>
<td>$X = 43$</td>
<td>$X = 46$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 16.5</td>
<td>s.d. = 15.8</td>
</tr>
<tr>
<td>Bio-Total</td>
<td>$X = 62.0$</td>
<td>$X = 61.8$</td>
</tr>
<tr>
<td></td>
<td>s.d. = 44.6</td>
<td>s.d. = 45.1</td>
</tr>
</tbody>
</table>
absenteeism measure. Consequently the differences between the personality and absenteeism correlations for the hold-out and development groups may be a function of sampling.

Hypothesis 1 was tested for each year that absenteeism data were collected. The experiment-wise error rate was set for each year at .01 (.05/5) to control for the number of correlations (Cohen & Cohen, 1983). The results of these analyses for all six years are presented in Table 22 on the following page. For 1991, the personality scale for Conscientiousness was significant for the developmental group \( (r = .21, p < .001) \), but not for the hold-out group \( (r = .04, p = .65) \). This result is most likely due to the initial differences between the hold-out and development samples for the Conscientiousness means reported earlier.

As expected, all of the biodata scored scales were significant for the developmental group \( (p < .01) \), and the Openness to Experience \( (r = .22, p = .004) \), Conscientiousness \( (r = .19, p = .009) \), and Total scales \( (r = .27, p < .001) \) were significant for the cross validation sample. Correlations for the biodata scored scales for Neuroticism \( (r = .18, p = .015) \) and Extraversion \( (r = .17, p = .02) \) were significant at the conventional level.

For the 1992 development group, the Conscientiousness personality scale was a significant predictor of absenteeism \( (r = .19, p < .001) \) and the Agreeableness personality scale approached significance \( (r = .09, p = .03) \) as a predictor. For the hold-out group, only the Conscientiousness personality scale approached significance \( (r = .11, p = .04) \). When biodata scoring was used, all of the scales were significant for the development group (correlations ranged from .22 for the Extraversion scale to .38 for the Total scale). For the cross validation group, all of the scales were also significant and the correlations ranged from .14 for Extraversion to .23 for the Total scale.

Similar results were obtained for the 1993 sample. Of the five personality scales, only the Conscientiousness scale was significant for either the development \( (r = .16, p < .001) \) or hold-out groups \( (r = .18, p < .001) \). Once again, all six biodata scored scales were significant for the development group with correlations ranging from .14 for Agreeableness to .34 for the Total scale. With the exception of the empirical Agreeableness scale \( (r = .07, p = .10) \), the biodata scored scales for
Table 22. Correlations Between Biodata and Personality Scales and Absenteeism.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>-0.13*</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>Extraversion</td>
<td>-0.03</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10*</td>
</tr>
<tr>
<td>Open/Exp</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.12*</td>
<td>0.10</td>
<td>0.04</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.00</td>
</tr>
<tr>
<td>Consc</td>
<td>0.04</td>
<td>0.21**</td>
<td>0.11*</td>
<td>0.18**</td>
<td>0.16**</td>
<td>0.08</td>
</tr>
<tr>
<td>Bio_N</td>
<td>0.18*</td>
<td>0.36**</td>
<td>0.15**</td>
<td>0.15**</td>
<td>0.20**</td>
<td>0.04</td>
</tr>
<tr>
<td>Bio_E</td>
<td>0.17*</td>
<td>0.28**</td>
<td>0.14**</td>
<td>0.22**</td>
<td>0.23**</td>
<td>0.19**</td>
</tr>
<tr>
<td>Bio_O</td>
<td>0.22**</td>
<td>0.34**</td>
<td>0.20**</td>
<td>0.24**</td>
<td>0.22**</td>
<td>0.19**</td>
</tr>
<tr>
<td>Bio_A</td>
<td>0.17*</td>
<td>0.34**</td>
<td>0.15**</td>
<td>0.07</td>
<td>0.14**</td>
<td>0.18**</td>
</tr>
<tr>
<td>Bio_C</td>
<td>0.19**</td>
<td>0.34**</td>
<td>0.19**</td>
<td>0.14**</td>
<td>0.26**</td>
<td>0.19**</td>
</tr>
<tr>
<td>Bio_Total</td>
<td>0.27**</td>
<td>0.45**</td>
<td>0.23**</td>
<td>0.23**</td>
<td>0.34**</td>
<td>0.23**</td>
</tr>
</tbody>
</table>

** p < .01, * p < .05
Neuroticism \((r = .15, p = .007)\), Extraversion \((r = .22, p < .001)\), Openness to Experience \((r = .24, p < .001)\), Conscientiousness \((r = .14, p = .01)\), and Total \((r = .23, p < .001)\) were significant for the cross validation sample.

For the 1994 sample, the Conscientiousness personality scale was significantly associated with absenteeism for the development group \((r = .15, p < .001)\), but not the hold-out group \((r = .08, p = .06)\). None of the other personality scales were significantly associated with absenteeism for either the development or hold-out groups. Significant correlations between the biodata scored scales and absenteeism for the development and hold-out groups were, respectively: Extraversion \((r = .20, p < .001; r = .19, p < .001)\), Openness to Experience \((r = .24, p < .001; r = .19, p < .001)\), Agreeableness \((r = .15, p < .001; r = .18, p < .001)\), Conscientiousness \((r = .20, p < .001; r = .19, p < .001)\), and Total \((r = .26, p < .001; r = .23, p < .001)\). Only the Neuroticism biodata scored scale was significant for the development group \((r = .13, p < .001)\), but not for the cross validation group \((r = .04, p = .22)\).

The personality scale for Conscientiousness was significant for the 1995 development \((r = .12, p < .001)\) and hold-out groups \((r = .14, p < .001)\), and the Extraversion personality scale \((r = .09, p = .003)\) was significant for the development group and approached significance for the hold-out group \((r = .10, p = .017)\). The scale for Agreeableness \((r = -.08, p = .009)\) was significant and Neuroticism \((r = .07, p = .02)\) approached significance for the development group; however, these scales were not significant when applied to the hold-out group. The 1995 biodata scored scales were all significant for the development group and the correlations ranged from a low of .18 for the Neuroticism scale to a high of .37 for the Total scale. With the exception of the Neuroticism biodata scored scale \((r = -.05, p = .14)\) all of the cross-validities were significant, with a range of .14 for Extraversion and Agreeableness to a high of .25 for the Conscientiousness scale \((p < .01)\).

Finally, the results of the 1996 sample were similar to those obtained in prior years. For the development sample, the correlations for the personality scales of Conscientiousness \((r = .13, p < .001)\), Agreeableness \((r = -.08, p = .008)\), and Extraversion \((r = .08, p = .004)\) were significant, and the Neuroticism scale approached significance \((r = .06, p = .04)\). When comparing the magnitude of the
correlations obtained in 1996 to those obtained in previous years it is evident that the significant results for Agreeableness and Extraversion are chiefly a function of sample size. None of the correlations for the personality scales were significant for the hold-out sample, and only the Conscientiousness scale approached significance \( (r = .09, p = .02) \). All of the biodata scored scales were significant for the 1996 development and hold-out groups \( (p < .01) \). The correlations for the development sample ranged from a low of .15 for the Neuroticism scale to a high of .31 for the Total scale. The shrinkage for the hold out sample was minimal and all of the biodata scored scales were significant with correlations ranging from .16 for the Neuroticism scale to .30 for the Total scale.

Taken as a whole, these results provide partial support for the hypothesis. For the development groups, the personality scale for Conscientiousness was significantly associated with absenteeism for all six years. However, for the hold-out sample the Conscientiousness scale was only significant for two of the six years (1993 and 1995), and approached significance in two others (1992 and 1996). Although not convincing, these findings do provide some support of the importance of Conscientiousness compared to other personality dimensions. Specifically, the other hypothesized scales of Neuroticism and Agreeableness did not demonstrate reliable associations with absenteeism. The correlation between Agreeableness and absenteeism was significant for three of the six years in the development sample but not the hold out sample and the magnitude of the correlations for the former sample was less than .10. Although the correlation between Neuroticism and absenteeism was highest for 1991 \( (r = -.13) \), the only significant correlations were obtained for the 1995 and 1996 development samples and the coefficients for the other years were consistently below .10.

The effectiveness of biodata scoring compared to personality scoring for the prediction of absenteeism was tested by examining each pair of correlations. Results of the \( Z \) (normal curve) test and the 95\% confidence intervals for each year are presented in Table 23. For the 1991 sample, the biodata scored scales for Extraversion \( (Z = 2.18, p = .01) \) and Conscientiousness \( (Z = 3.02, p = .001) \) significantly outperformed their personality counterparts. None of the other correlations differed significantly.
The result for the superiority of the biodata scored Conscientiousness scale was not replicated for the 1992 sample ($Z = 1.68$, $p = .04$), nor was the result for Extraversion ($Z = .82$, $p = .44$). Instead, the scales for Neuroticism ($Z = 2.71$, $p = .002$) differed for the 1992 sample, and Openness to Experience ($Z = 1.90$, $p = .02$) approached significance with the biodata scored scales being better predictors in both sets of comparisons.

For the 1993 sample, only the biodata scored scales for Extraversion ($Z = 3.51$, $p < .001$) and Openness to Experience ($Z = 2.92$, $p = .002$) were superior to the personality predictors. The 1993 Conscientiousness biodata scored scale did not outperform the personality scale.

The biodata scored scales for Extraversion ($Z = 3.20$, $p = .004$) outperformed the personality scales for the 1994 sample. The biodata scored scales for Openness to Experience ($Z = 1.90$, $p = .02$), Agreeableness ($Z = 2.17$, $p = .015$), and Conscientiousness ($Z = 2.70$, $p = .002$) approached significance as being better predictors of absenteeism.

For the 1995 sample, only the Openness to Experience ($Z = 4.11$, $p < .001$) and Agreeableness ($Z = 2.92$, $p = .002$) biodata scored scales outperformed the personality scales. By contrast, the entire set of biodata scored scales significantly outperformed the personality scales for the 1996 sample.
Differences between the personality and biodata scored scales for Conscientiousness ($Z = 5.10$, $p < .001$), Neuroticism ($Z = 4.83$, $p < .001$), Extraversion ($Z = 4.03$, $p < .001$), Openness to Experience ($Z = 3.65$, $p < .001$), and Agreeableness ($Z = 3.27$, $p < .001$) demonstrated the superiority of the empirical scored scales.

Hypothesis 2 stated that the empirically keyed scales would provide incremental variance above and beyond that of the personality scales. Consistent with previous analyses, the six scales of the CPI were entered in a stepwise regression followed by the six empirical scales in a stepwise block. This analysis was conducted separately for each year that absenteeism data were collected.

For the 1991 sample, none of the personality measures were significant ($p < .05$) when entered in the first block. Only the Neuroticism scale approached significance [$F(3, 171) = 2.76$, $p = .09$] and was retained in the equation ($R^2 = .02$). When the biodata scored scales were entered as a stepwise block, the biodata scored scales for Neuroticism and Openness to Experience were added to the equation. The addition of these two scales increased the $R^2$ to .09 [$F(3, 171) = 5.87$, $p < .001$].

As a supplementary analysis the set of ten scales were simultaneously entered into a regression equation for each of the six years. The supplementary analysis was completed for the six years that data were collected and the resulting $R^2$ and beta weight for each personality and biodata scored scale are presented in Table 24 on the following page. For the 1991 sample, the resulting $R^2$ of .15 when all ten scales were entered was significant [$F(10, 164) = 2.83$, $p = .003$]. An examination of the beta weights revealed that the Conscientiousness ($\beta = -.30$, $p = .01$) personality scale, the Openness to Experience biodata scored scale ($\beta = .20$, $p = .03$), and the Neuroticism biodata scored scale ($\beta = .19$, $p = .05$) were significant.

For the 1992 sample, none of the personality scales entered the stepwise equation ($p > .05$) for the first block. In the second stepwise entry, only the biodata scored scale for Openness to Experience was significant [$F(1, 270) = 10.76$, $p = .001$] yielding a $R^2$ of .04. In contrast, when the set of ten scales was entered into the equation the $R^2$ increased to .07 [$F(10, 261) = .033$]; however, none of the beta weights for the personality or biodata scored scales were significant.
Table 24. Regression Results for CPI Personality and Biodata Scales Entered as a Simultaneous Block.

<table>
<thead>
<tr>
<th>Scale</th>
<th>1991 (n = 175)</th>
<th>1992 (n = 272)</th>
<th>1993 (n = 274)</th>
<th>1994 (n = 396)</th>
<th>1995 (n = 493)</th>
<th>1996 (n = 660)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near</td>
<td>p = -.10</td>
<td>SigT = .34</td>
<td>SigT = -.01</td>
<td>SigT = -.05</td>
<td>SigT = -.16</td>
<td>SigT = -.14</td>
</tr>
<tr>
<td>Extra</td>
<td>p = -.04</td>
<td>SigT = .68</td>
<td>SigT = .07</td>
<td>SigT = .35</td>
<td>SigT = -.14</td>
<td>SigT = .05</td>
</tr>
<tr>
<td>Open</td>
<td>p = .05</td>
<td>SigT = .59</td>
<td>SigT = .01</td>
<td>SigT = -.10</td>
<td>SigT = .15</td>
<td>SigT = .09</td>
</tr>
<tr>
<td>Agree</td>
<td>p = 1.0</td>
<td>SigT = .34</td>
<td>SigT = .07</td>
<td>SigT = .45</td>
<td>SigT = .28</td>
<td>SigT = .01</td>
</tr>
<tr>
<td>Consc</td>
<td>p = -.30</td>
<td>SigT = -.12</td>
<td>SigT = .20</td>
<td>SigT = .06</td>
<td>SigT = .48</td>
<td>SigT = -.22</td>
</tr>
<tr>
<td>Bio_N</td>
<td>p = .19</td>
<td>SigT = .05</td>
<td>SigT = .10</td>
<td>SigT = -.16</td>
<td>SigT = .04</td>
<td>SigT = .23</td>
</tr>
<tr>
<td>Bio_E</td>
<td>p = .07</td>
<td>SigT = .39</td>
<td>SigT = .02</td>
<td>SigT = .74</td>
<td>SigT = .22</td>
<td>SigT = .02</td>
</tr>
<tr>
<td>Bio_O</td>
<td>p = .20</td>
<td>SigT = .20</td>
<td>SigT = .15</td>
<td>SigT = .86</td>
<td>SigT = .23</td>
<td>SigT = .21</td>
</tr>
<tr>
<td>Bio_A</td>
<td>p = .02</td>
<td>SigT = .79</td>
<td>SigT = .00</td>
<td>SigT = .90</td>
<td>SigT = .28</td>
<td>SigT = .01</td>
</tr>
<tr>
<td>Bio_A</td>
<td>p = .23</td>
<td>SigT = .08</td>
<td>SigT = .15</td>
<td>SigT = -.11</td>
<td>SigT = .82</td>
<td>SigT = .28</td>
</tr>
</tbody>
</table>

The Conscientiousness personality scale was significant in the first stepwise block for the 1993 sample [F(1, 270) = 9.08, p = .003] producing a R² of .03. The second stepwise equation included the biodata scored scales for Extraversion and Openness to Experience scales that increased the R² to .09 [F(3, 270) = 8.44, p < .001]. In comparison, entering the ten scales simultaneously to the equation resulted in a R² of .16 [F(10, 263) = 4.85, p < .001]. The beta weights for the Agreeableness personality scale (β = .28, p = .006), and the empirical scales for Neuroticism (β = .16, p = .04), Extraversion (β = .22, p = .02), and Openness to Experience (β = .23, p = .004) were significant.

None of the personality scales entered the equation in the first stepwise block (p > .05) for the 1994 absenteeism sample. However, for the second block the inclusion of the biodata scored scales for Extraversion and Openness to Experience resulted in a R² of .06 [F(2, 393) = 12.17, p < .001]. Entering the ten scales as a block produced a R² of .12 [F(10, 385) = 5.22, p < .001]. The beta weights for the personality scales for Extraversion (β = -.16, p = .05) and Conscientiousness (β = -.22, p = .002) were significant, as were the biodata scored scales for Extraversion (β = .21, p = .02), Agreeableness (β = .19, p = .01), and Conscientiousness (β = .28, p = .001).
The personality scale for Conscientiousness was the only scale to enter in the first stepwise block for the 1995 sample yielding a $R^2$ of .02 [$F(1, 490) = 10.36, p = .001$]. The biodata scored scale for Conscientiousness was included in the second block of the stepwise that improved the $R^2$ to .07 [$F(2, 490) = 17.09, p < .001$]. Entering the ten scales resulted in an improved $R^2$ of .32 [$F(10, 482) = 5.68, p < .001$]. Neuroticism ($\beta = -.23, p = .03$), Agreeableness ($\beta = .14, p = .03$), and Conscientiousness ($\beta = -.18, p = .01$) had significant beta weights for the personality scales, and Openness to Experience ($\beta = -.12, p = .04$) and Conscientiousness ($\beta = -.31, p < .001$) had significant beta weights for the biodata scored scales.

Finally, for the 1996 sample the Conscientiousness personality scale entered the equation in the first stepwise block [$F(1, 656) = 5.35, p = .02$] producing a $R^2$ of .01. Once again the addition of biodata scored scales improved prediction. Here the inclusion of the Conscientiousness and Extraversion biodata scored scales improved the $R^2$ to .10 [$F(3, 656) = 23.07, p < .001$]. Simultaneously entering the ten scales as a block resulted in a $R^2$ of .12 [$F(10, 649) = 8.73, p < .001$]. Beta weights for the personality scales for Neuroticism ($\beta = -.14, p = .009$) and Conscientiousness ($\beta = -.19, p < .001$) were significant, and the biodata scored scales for Neuroticism ($\beta = .11, p = .03$), Extraversion ($\beta = .13, p = .005$), and Conscientiousness ($\beta = .26, p < .001$) were significant.

Taken together, the findings reported above support hypothesis 2. In every case the biodata scored scales provided incremental variance above and beyond that of the personality scales. Nonetheless, some inconsistencies are noted when comparing the results from the six years of data. Specifically, whereas the personality scale for Conscientiousness was the only scale to be included in the final equation, it did so in only three of the six years. Further inconsistencies were noted in the specific biodata scored scales that provided incremental variance above the personality scales. For the years 1991 through 1994 the scale for Openness to Experience provided incremental variance beyond the other scales. However, during this same period the Neuroticism scale was only included in 1991, and the Extraversion biodata scored scale was only included in 1993 and 1994. For 1995 and 1996, the biodata scored scale for Conscientiousness provided incremental variance over the personality measure.
even though the personality scale for Conscientiousness was already in the equation. The two years differed such that the 1996 equation included the biodata scored scale for Extraversion that was not present in the 1995 equation, but had previously been included for the 1994 and 1993 samples.

Hypothesis 3 compared the personality and biodata scored PAMS profiles for the high and low absenteeism groups and was analyzed separately for each year of absenteeism data. Table 25 shows the

<table>
<thead>
<tr>
<th>Table 25. PAMS Dimension Values for High and Low Absenteeism Groups.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1991</td>
</tr>
<tr>
<td>Near</td>
</tr>
<tr>
<td>Extra</td>
</tr>
<tr>
<td>Open</td>
</tr>
<tr>
<td>Agree</td>
</tr>
<tr>
<td>Consc</td>
</tr>
<tr>
<td>Bio_N</td>
</tr>
<tr>
<td>Bio_O</td>
</tr>
<tr>
<td>Bio_E</td>
</tr>
<tr>
<td>Bio_A</td>
</tr>
<tr>
<td>Bio_C</td>
</tr>
<tr>
<td>1992</td>
</tr>
<tr>
<td>Near</td>
</tr>
<tr>
<td>Extra</td>
</tr>
<tr>
<td>Open</td>
</tr>
<tr>
<td>Agree</td>
</tr>
<tr>
<td>Consc</td>
</tr>
<tr>
<td>Bio_N</td>
</tr>
<tr>
<td>Bio_O</td>
</tr>
<tr>
<td>Bio_E</td>
</tr>
<tr>
<td>Bio_A</td>
</tr>
<tr>
<td>Bio_C</td>
</tr>
</tbody>
</table>
| Note: * Only absolute value elevations greater than one were used to interpret profiles.
dimension values for the high and low absenteeism groups for each year. For all six years, a two-
dimensional model provided a good fit to the data (STRESS < .06). Hence the high and low groups
will be compared for dimension one and dimension two for the empirical and personality scales.

The 1991 sample produced a similar personality profile for the first dimension for the high and
low absenteeism groups. Both groups' dimensions were characterized by a positive elevation for
Neuroticism and a negative elevation for Agreeableness. In other words, being well adjusted and being
unwilling to get along with others may describe one profile evident in law enforcement occupations for
either a high or low absenteeism individual. The two groups did differ in the composition of the
second personality dimension. Here, the high absenteeism profile contained a negative elevation for
Openness to Experience, whereas the low absenteeism group contained a negative elevation for
Conscientiousness. As such, the high absenteeism profile may be defined as not being interested in
cultural pursuits, and the low absenteeism profile may be defined as being low in work ethic and
reliability.

Differences for the two dimensions for the high and low absenteeism groups were more dramatic
for the biodata scored dimensions. The high absenteeism group had a negative elevation for Openness
to Experience coupled with a positive elevation for the Agreeableness scale. In contrast the low
absenteeism group had a positive elevation for the Openness to Experience scale and a negative
elevation for the Agreeableness scale. For dimension two, the high absenteeism profile contained a
positive elevation for Neuroticism and a high negative elevation for Extraversion; whereas the low
absenteeism profile was characterized by a negative elevation for the Neuroticism biodata scored scale
and a high positive elevation for the Extraversion scale.

High and low absenteeism group profiles also differed for the personality scales for the 1992
sample. For the first dimension, the high absenteeism group was characterized by a positive elevation
for the Extraversion scale with a negative elevation for the Agreeableness scale. In other words,
extraverted people, who are less easy-going and able to get along with others may best describe the first
dimension. On the other hand, the low absenteeism group had a profile containing a large elevation
for Neuroticism and a large negative elevation for Agreeableness. This profile might be described as being well adjusted, less easygoing, and less able to get along with others. For dimension two the high absenteeism profile was driven by a positive elevation for Openness to Experience scales with a large negative elevation for Neuroticism scale. The high absenteeism profile might be defined as interested in intellectual and cultural pursuits and not being well adjusted. By comparison, the low absenteeism profile that had a negative elevation for Conscientiousness, described as not having a well-developed work ethic.

Biodata scoring again altered the profiles for both groups' profile dimensions. The high absenteeism profile for dimension one contained a positive elevation for Neuroticism and a negative elevation for Agreeableness; whereas the low absenteeism group contained a large negative elevation for Openness to Experience and a positive elevation for Conscientiousness. Dimension two for the high absenteeism profile contained no meaningful elevations and the low absenteeism group contained a positive elevation for Neuroticism combined with a negative elevation for Extraversion.

The personality profiles derived for the 1993 sample continued the trend for the Agreeableness scale obtained in previous samples. Moreover, the results for the 1993 sample paralleled those obtained in the 1991 sample. Here, the high and low absenteeism profiles contained a positive elevation for Neuroticism and a negative elevation for Agreeableness for the first profile dimension. The prototypical profile for both high and low absenteeism may be described as one who is well adjusted and does not get along well with others. For the second profile dimension, a negative elevation for the Openness to Experience scale drove the profile for the high absenteeism group, and a positive elevation for the Conscientiousness scale defined the profile for the low absenteeism group. Therefore, the high absenteeism profile reflects those who have an interest in cultural and artistic pursuits, whereas the low absenteeism profile represents those who have a strong work ethic and are conscientiousness.

The two biodata scored profile dimensions differed from those derived for the personality profiles. For the first dimension, a positive elevation for the Openness to Experience was evident in the high
absenteeism biodata scored profile, whereas a negative elevation for Openness to Experience and a positive elevation for Agreeableness was present in the low absenteeism group. As such, the high absenteeism group profile is likely to be interested in cultural and intellectual pursuits, whereas the low absenteeism profile is not likely to be interested in these pursuits and is likely to be friendly with others. Dimension two for the high absenteeism group had a positive elevation for the Neuroticism scale, and the low absenteeism profile dimension had a positive elevation for the Conscientiousness biodata scored scale. Therefore, the high absenteeism group profile is characterized by adjusted individuals, whereas the low absenteeism profile is defined by work ethic and conscientiousness.

The first personality dimension profile derived for the 1994 sample was similar to the 1993 for the low and high absenteeism groups. Positive elevations for the high and low absenteeism groups were obtained for the Neuroticism scale with large positive elevations for the Agreeableness scale. When comparing the second dimension profiles, the high absenteeism group did not contain notable elevations, but the low absenteeism profile contained a negative elevation for Conscientiousness. Consequently, and contrary to prediction, the prototypical profile for low absenteeism group is one in which individuals are low in work ethic or conscientiousness. For the 1994 biodata scored scales, the first dimension profile exhibited some similarities such that both groups contained a positive elevation for Neuroticism and a negative elevation for Openness to Experience. The high and low absenteeism groups differed because the low absenteeism profile contained a positive elevation for Agreeableness, not present in the high absenteeism group. The biodata scored profiles for dimension two were also similar because both absenteeism groups produced profiles that contained a positive elevation for Neuroticism.

The profiles derived for 1995 support results obtained for previous years. For the first dimension, both profiles contained a positive elevation for Neuroticism and a negative elevation for Agreeableness, described as being well adjusted and not being able to get along well with others. For 1995, the dimension two profiles were also similar as both absenteeism groups contained a positive elevation for Conscientiousness, described as having a good work ethic.
The 1995 biodata scored profiles were also similar to the 1994 profiles because both have a positive elevation for the Neuroticism scale and a negative elevation for the Openness to Experience scale for the low and high absenteeism groups for dimension one. For the biodata scored profile dimension two, the high absenteeism profile was defined by a positive elevation for Neuroticism, but the low absenteeism profile was driven by a negative elevation for the biodata scored Conscientiousness scale.

Finally, the low and high absenteeism 1996 personality profiles for dimension one contained a positive elevation for Neuroticism and a negative elevation for Agreeableness that is consistent with previous years. Differences between the high and low absenteeism groups are most notable for dimension two with the high absenteeism group having a profile with no meaningful elevations, whereas the low absenteeism profile had a positive Conscientiousness driven profile.

The 1996 biodata scored scales for the absenteeism groups were similar with respect to positive Neuroticism and negative Openness to Experience scales for dimension one. The only slight difference in the groups’ profiles was noted for dimension two because the high absenteeism group had a positive elevation for Neuroticism and a negative elevation for Extraversion, and the low absenteeism group had a negative elevation for Extraversion but not a positive elevation for Neuroticism.

To summarize, the dimension one profile for the personality scales for both the high and low absenteeism groups consistently contained a positive elevation for Neuroticism and a negative elevation for Agreeableness with the exception of the high absenteeism group in 1992. A second consistency was noted in the low absenteeism group in which the Conscientiousness scale defined the second profile for all six years. By contrast, the Conscientiousness scale was elevated only for the 1995 sample for the high absenteeism group. A second difference between groups was noted such that the Openness to Experience elevation defined the second dimension of the high absenteeism profile for the first three years but not years 1994 through 1996.

When comparing the biodata scored profiles no notable trends emerge. As noted before, the composition of scales changed somewhat because only the items that best predicted absenteeism for
that particular year were included in the biodata scored scales. Therefore, by altering the scoring, it is also likely that the construct being measured by the scale may have changed, as well as the relationship of the scales to each other. Because the scales may have changed in meaning it may not be prudent to expect there to be similarities across years.

Despite the consistencies and trends noted above, none of the results directly supported the hypothesis that the biodata scored profiles would contain larger elevations for Conscientiousness, Agreeableness, and Neuroticism than the traditionally scored scales. Instead, results suggest that biodata scoring not only may change the meaning of the scales but also the meaning of the profiles. Bear in mind that all of the biodata scored scales did not contain the entire set of items designed to measure the associated personality scales and there was a reduction in the reliability of the derived biodata scored scales. Therefore, the interpretation of the biodata scored scale as a valid measure of a Big Five personality construct may be suspect.

Hypothesis 5 which proposed that the biodata scored level parameter would be a better predictor of absenteeism than the personality scored level parameter was also tested for each year. Only hold-out groups were used for the analyses. For the 1991 sample, the personality level parameter (X = -.031, s.d. = .46) produced a R^2 of .00 that was not significant [F(1, 173) = .02, p = .89]. However, as predicted, the biodata scored level parameter (X= -.02, s.d. = .75) increased the R^2 significantly to .07 [F(2, 172) = 6.55, p = .002].

Results were similar for the 1992 sample as the personality level parameter (X = -.02, s.d. = .45) had a R^2 of .01 that was not significant [F(1, 269) = 2.42, p = .12], but entering the biodata scored level parameter (X = -.03, s.d. = .71) significantly increased the R^2 to .05.

The level parameter for biodata scoring was also a better predictor of absenteeism for the 1993 sample. Here, the R^2 of .01 for the personality level parameter (X = .00, s.d. = .45) was not significant [F(1, 272) = 3.40, p = .07], but the R^2 of .06 for the biodata scored level parameter (X = .01, s.d. = .71) was significant [F(1, 272) = 7.94, p < .001].
Results from the 1994 sample replicated the findings reported above. Here, the $R^2$ of .01 was not significant [$F(1, 394) = 2.24, p = .14$] for the personality level parameter ($\bar{X} = .00, s.d. = .46$), but increased to .05 [$F(2, 393) = 10.24, p < .001$] when the biodata scoring level parameter was entered into the equation ($\bar{X} = -.01, s.d. = .72$).

1995 results provide additional support for the hypothesis. The personality level parameter ($\bar{X} = .04, s.d. = .44$) produced a $R^2$ of .01 that was not significant [$F(1, 491) = 2.58, p = .11$], whereas entering the biodata scoring level parameter ($\bar{X} = .01, s.d. = .67$) increased the $R^2$ to .05 [$F(2, 490) = 12.16, p < .001$].

Finally, the 1996 results continued the trend reported above. Entering the personality level parameter ($\bar{X} = -.01, s.d. = .43$) at the first step ($R^2 = .00$) was not significant [$F(1, 658) = .00, p = .95$], but entering the biodata scoring level parameter ($\bar{X} = -.01, s.d. = .71$) increased the $R^2$ to .08 [$F(2, 657) = 30.39, p < .001$].

The findings above strongly support the hypothesis. In every case, the scoring level parameter provided incremental variance above and beyond that of the personality level parameter.

Hypothesis 6 compared the mean correspondence weights of the high absenteeism group to the weights of the low absenteeism group. These analyses were also conducted for each year that absenteeism data were collected. For 1991, the high ($\bar{X} = -.04, s.d. = .52$) and low ($\bar{X} = .06, s.d. = .45$) groups did not differ in the mean correspondence weights for personality profile one [$F(1, 113) = .99, p = .32$]. Also, the absenteeism group means for the high ($\bar{X} = .08, s.d. = .71$) and low ($\bar{X} = .01, p = .61$) absenteeism groups did not differ significantly for the second profile dimension [$F(1, 113) = .34, p = .56$]. The high and low groups did not differ significantly for the two biodata scored dimensions. For the first dimension, the group means for the high and low respectively were -.03 (s.d. = .04) and .04 (s.d. = .31) and these means did not differ [$F(1, 113) = 1.25, p = .27$]. Similarly for the second profile dimension the high ($\bar{X} = .01, s.d. = .49$) and low means ($\bar{X} = .01, s.d. = .39$) did not significantly diverge [$F (1, 113) = .004, p = .95$].
For 1992, the mean correspondence weights for the high ($\bar{X} = -.01$, s.d. = .53) and low ($\bar{X} = -.02$, s.d. = .48) absenteeism groups did not differ for the first personality dimension [$F(1, 188) = .013$, $p = .91$], but did for the second [$F(1,188) = 4.38$, $p = .04$] with means of .07 (s.d. = .07) and -.15 (s.d. = .76), respectively. For the biodata scoring dimensions there were no significant differences for either of the mean correspondence weights. For dimension one, the correspondence weights for the high ($\bar{X} = .05$, s.d. = .38) and low absenteeism groups ($\bar{X} = -.02$, s.d. = .31) did not differ [$F(1, 188) = 1.78$, $p = .18$], nor did the high ($\bar{X} = .02$, s.d. = .45) and low ($\bar{X} = .06$, s.d. = .44) means differ for dimension two.

For the 1993 sample none of the mean correspondence weights for personality or biodata scoring profiles differed significantly. The high ($\bar{X} = .04$, s.d. = .47) and low ($\bar{X} = .03$, s.d. = .44) means did not differ for the personality dimension [$F (1, 178) = .03$, $p = .85$], nor did the high ($\bar{X} = -.01$, s.d. = .67) and low ($\bar{X} = -.12$, s.d. = .58) groups differ for dimension two. For the first biodata scoring dimension, the high ($\bar{X} = .03$, s.d. = .34) and low ($\bar{X} = .03$, s.d. = .34) means did not significantly differ [$F (1, 178) = .01$, $p = .91$], nor did the respective means of .05 (s.d. = .95) and .13 (s.d. = 1.03) differ for the second biodata scoring dimension [$F(1, 178) = .31$, $p = .58$].

Similarly the high and low absenteeism groups from the 1994 sample did not have significantly different mean correspondence weights. Mean weights for personality dimension one for high and low absenteeism were -.01 (s.d. = .45) and .05 (s.d. = .47), respectively and these were not significantly different [$F(1, 265) = .97$, $p = .33$]. Likewise the mean weights for personality dimension two were -.04 (s.d. = .67) for the high and -.01 (s.d. = .85) for the low which did not differ significantly [$F(1, 265) = .15$, $p = .70$]. The two dimensions for the biodata scoring profiles also did not differ as a function of absenteeism group. For biodata scoring dimension one, the high mean of -.02 (s.d. = .35) and low mean of .02 (s.d. = .36) did not differ [$F(1, 265) = .76$, $p = .38$], and for dimension two the high mean of .03 (s.d. = .55) and low mean of .06 (s.d. = .60) also did not differ significantly [$F(1, 265) = .18$, $p = .67$].
For the 1995 sample, the mean correspondence weights for the high ($\bar{X} = -.04$, s.d. = .48) and low ($\bar{X} = -.02$, s.d. = .48) absenteeism groups did not differ for dimension one [$F(1, 308) = .10$, $p = .75$], nor dimension two [high $\bar{X} = .00$, s.d. = .71; low $\bar{X} = .07$, s.d. = .75; $F(1, 308) = .45$, $p = .50$] for the personality dimensions. Similarly, the mean correspondence weights for the high ($\bar{X} = .00$, s.d. = .32) and low ($\bar{X} = .07$, s.d. = .36) absenteeism groups did not differ for dimension one [$F(1, 308) = .45$, $p = .50$], nor dimension two [high $\bar{X} = .03$, s.d. = .49; low $\bar{X} = -.03$, s.d. = .52; $F(1, 308) = 1.23$, $p = .27$]

Finally, the results for the 1996 sample continued the trend of nonsignificant findings. For personality dimension one, the mean correspondence weights for the high ($\bar{X} = .00$, s.d. = .50) and low ($\bar{X} = .03$, s.d. = .44) absenteeism groups did not significantly differ [$F(1, 477) = .28$, $p = .60$], and for dimension two the mean correspondence weights for the high ($\bar{X} = -.03$, s.d. = .81) and low ($\bar{X} = -.07$, s.d. = .71) absenteeism groups also did not differ [$F(1, 477) = .34$, $p = .56$]. The mean correspondence weights for the biodata scoring dimensions one and two for the high absenteeism group were .00 (s.d. = .35) and .02 (s.d. = .53), and for the low absenteeism group the mean correspondence weights were -.01 (s.d. = .35) and -.02 (s.d. = .53). The high and low absenteeism means did not differ for dimension one [$F(1, 477) = .11$, $p = .75$], nor dimension two [$F(1, 477) = .78$, $p = .38$]. Based on the total findings above, the hypothesis that the mean correspondence weights for the high and low absenteeism groups would differ significantly was not supported for either dimension of the personality or biodata scoring profiles.

Hypothesis 7 provided an additional test of the importance of Conscientiousness. Here, the biodata scored scale for Conscientiousness was compared to the biodata scored scales for Neuroticism and Agreeableness. The correlations were tested by assigning the biodata scored Conscientiousness scale a weight of “1,” the Neuroticism and Agreeableness scales a weight of “-.5,” and the remaining scales a weight of “0.” This comparison was completed for each year. For 1991, the Conscientiousness biodata scored scale ($r = .19$) was not a better predictor than Neuroticism ($r = .18$), or Agreeableness ($r = .17$) as evidenced by the $Z$ of .21. Similarly, the comparison of scales for 1992 showed that Conscientiousness scale ($r = .19$) was not a more effective predictor than Neuroticism ($r = .15$) or
Agreeableness ($r = .15; Z = .22, p = .91$). Correlations for the three scales in 1993 were
Conscientiousness ($r = .14$), Agreeableness ($r = .07$), and Neuroticism ($r = .15$) and this a priori contrast was also not significant ($Z = .55, p = .79$). The comparison for 1994, however, did approach significance ($Z = 1.63, p = .051$). For this year the Conscientiousness ($r = .19$) scale was a better predictor than the Neuroticism ($r = .04$) and Agreeableness ($r = .18$) scales due to the low correlation of the former. In 1995, the Conscientiousness scale ($r = .25$) was significantly better ($Z = 3.62, p = .002$) than either the Neuroticism ($r = .05$) or Agreeableness ($r = .14$) scales. Similarly, the a priori contrast was significant for 1996 ($Z = 1.88, p = .03$) with Conscientiousness ($r = .25$) being a better predictor than Neuroticism ($r = .16$) or Agreeableness ($r = .19$). In short, the a priori contrast were only significant for two of the six years and hence the results above do not provide unequivocal support for the hypothesis. These findings are primarily a result of the improvement that biodata scoring had on all of the scales, regardless of the initial correlation between the personality scale and absenteeism.

As a further test of the importance of the Conscientiousness scale, its item weights were compared to the item weights of the other scales and it was hypothesized that the contribution of the Conscientiousness scales would be greater, as reflected by the mean item weights, than the other two scales. As before, item weights were converted to absolute values to control for positive and negative values.

Prior to conducting the a priori contrasts, a repeated measures Anova was conducted to determine if the weighting of items changed across the years that data were collected. For this analysis the 462 items of the CPI were considered subjects, the five scales of the CPI comprised the between subject effect (scale type), the measures (i.e., items) were repeated for the six years that data were collected (Years), and the dependent variable was the weight assigned to each item. The overall effect for the between subject effect of Scale Type approached significance ($F(4, 457) = 2.35, p = .053$), the Year within subject effect was significant ($S = 1, M = 1 1/2, N = 225 1/2); F = 5.44, p < .001$), and the Scale Type by Year interaction was not significant ($S = 4, M = 0, N = 225 1/2); F = 1.09, p = .35$). These
results indicate that the item weights and the items that contributed to the scales changed as a function of year.

Follow-up univariate Anovas were conducted for each year to compare the importance of the Conscientiousness biodata scored scale. The mean, standard deviation, and the results for the univariate Anovas are presented in Table 26. Also presented are the results for the a priori contrasts.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>.89</td>
<td>.85</td>
<td>.73</td>
<td>.84</td>
<td>.67</td>
<td>.74</td>
</tr>
<tr>
<td>Extra</td>
<td>.80</td>
<td>.81</td>
<td>.95</td>
<td>.87</td>
<td>.91</td>
<td>.92</td>
</tr>
<tr>
<td>Open</td>
<td>.72</td>
<td>.86</td>
<td>.90</td>
<td>.88</td>
<td>.73</td>
<td>.87</td>
</tr>
<tr>
<td>Agree</td>
<td>.77</td>
<td>.90</td>
<td>.63</td>
<td>.77</td>
<td>.67</td>
<td>.87</td>
</tr>
<tr>
<td>Cons</td>
<td>.92</td>
<td>.89</td>
<td>.94</td>
<td>.93</td>
<td>.74</td>
<td>.87</td>
</tr>
<tr>
<td>F(4,457)</td>
<td>.86, p = .49</td>
<td>2.45, p = .05</td>
<td>1.13, p = .34</td>
<td>1.79, p = .13</td>
<td>.98, p = .42</td>
<td>2.11, p = .08</td>
</tr>
</tbody>
</table>

A Priori t(457) .871, p = .38 | 2.48, p = .01 | .739, p = .46 | .742, p = .46 | .646, p = .32 | 1.688, p = .092 |

whereby the Conscientiousness (weight of 1) biodata scored scale was compared to the Neuroticism (weight of -.5) and Agreeableness (weight of -.5) biodata scored scales. With the exception of the 1992 sample [F(4,457) = 2.45, p = .05] none of the other univariate Anovas were significant (p > .08).

Similarly, none of the a priori contrasts were significant except for 1992 [t(457) = 2.48, p = .01] with the Conscientiousness mean (X = .94, s.d. = .93) being greater than the combination of means for Neuroticism (X = .73, s.d. = .84) and Agreeableness (X = .63, s.d. = .77). Therefore, this hypothesis was not supported as the mean weights for the five scales did not differ significantly and Conscientiousness items were not assigned greater weights.

For Sample Three, the predictive validity of personality and biodata scored scales were assessed using a turnover criterion. This analysis was conducted for three groups. The first group consisted of the total sample of law enforcement employees including clerical and administrative support employees. The second and third groups consisted of law enforcement officers from a sheriff...
department and a probation department. Empirical scales were developed separately for each group.

The Cronbach’s alphas for the personality scale, the biodata scored scale, and the percentage of items used to develop the biodata scored scales are presented in Table 27. As shown, the reliabilities of the

<table>
<thead>
<tr>
<th>Scale</th>
<th>Aggregate</th>
<th>Probation</th>
<th>Sheriff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neur</td>
<td>.86</td>
<td>.40</td>
<td>32.8</td>
</tr>
<tr>
<td>Extra</td>
<td>.81</td>
<td>.45</td>
<td>44.7</td>
</tr>
<tr>
<td>Open</td>
<td>.69</td>
<td>.41</td>
<td>42.3</td>
</tr>
<tr>
<td>Agree</td>
<td>.75</td>
<td>.31</td>
<td>38.0</td>
</tr>
<tr>
<td>Consc</td>
<td>.81</td>
<td>.72</td>
<td>36.0</td>
</tr>
</tbody>
</table>

biodata scored scales differed from the personality reliabilities. In addition, the percentage of personality items used to develop the biodata scored scales varied from sample to sample. These results contribute to the previous conclusion that biodata scored scales are likely measuring multidimensional constructs different from those measured via personality scoring.

Hypothesis 9 predicted that a significant relationship would exist between turnover and Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability and these relationships would be greater using empirical rather than traditional scoring. Correlations between the five scales for the traditionally scored CPI and the six scales for biodata scoring, that includes a total biodata scored scale, are presented in Table 28 for the three groups specified above. None of the correlations between the personality scales and turnover were significant for the aggregate hold-out or development samples (p > .10), and all of the reported correlations were below .10. For the sheriff’s sample the scales for Openness to Experience (r = -.10, p = .01), Extraversion (r = -.12, p = .01), and Conscientiousness (r = .10, p = .01) were significantly different than zero for the development sample, but not for the hold-out sample. None of the other scales were significant for either the hold-out or
Table 28. Correlations Between Personality and Biodata Scales with Turnover by Occupation.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Aggregate Sample</th>
<th>Sheriff</th>
<th>Probation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cross (n = 555)</td>
<td>Dev (n = 980)</td>
<td>Cross (n = 308)</td>
</tr>
<tr>
<td>Neur</td>
<td>-.02</td>
<td>-.01</td>
<td>-.01</td>
</tr>
<tr>
<td>Open/Exp</td>
<td>.03</td>
<td>-.03</td>
<td>.01</td>
</tr>
<tr>
<td>Extra</td>
<td>-.04</td>
<td>.01</td>
<td>-.01</td>
</tr>
<tr>
<td>Agree</td>
<td>-.00</td>
<td>-.04</td>
<td>-.02</td>
</tr>
<tr>
<td>Consc</td>
<td>.08</td>
<td>.09</td>
<td>.02</td>
</tr>
<tr>
<td>Bio_N</td>
<td>-.14**</td>
<td>-.20**</td>
<td>-.19**</td>
</tr>
<tr>
<td>Bio_O</td>
<td>-.21**</td>
<td>-.24**</td>
<td>-.19**</td>
</tr>
<tr>
<td>Bio_E</td>
<td>-.14**</td>
<td>-.18**</td>
<td>-.13*</td>
</tr>
<tr>
<td>Bio_A</td>
<td>-.19**</td>
<td>-.19**</td>
<td>-.12*</td>
</tr>
<tr>
<td>Bio_C</td>
<td>-.12**</td>
<td>-.17**</td>
<td>-.17**</td>
</tr>
<tr>
<td>Bio_Total</td>
<td>-.23**</td>
<td>-.28**</td>
<td>-.26**</td>
</tr>
</tbody>
</table>

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

development groups (p > .10). Finally, for the probation department, the Neuroticism (r = -.13, p = .03) and Extraversion (r = .14, p = .02) scales approached significance at the .01 level the development sample, but not for the hold-out sample (p > .10).

Scoring the CPI scales using biodata methodology resulted in all of the scales being significant for the development samples for the aggregate, sheriff, and probation samples. The correlations for the aggregate sample ranged from -.18 (p < .001) for the Extraversion biodata scored scale to -.28 (p < .001) for the Total biodata scored scale. For the sheriff sample, correlations ranged from a low of -.18 for the Agreeableness biodata scored scale (p < .001) to a high of -.34 (p < .001) for the Total biodata scored scale. Finally the correlations for the probation department were greater than those obtained for the other groups with a range of -.31 for the biodata scored scales of Neuroticism, Agreeableness, and Conscientiousness (p < .001) to a high of -.47 (p < .001) for the Total biodata scored scale.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
More important, were the significant findings for the cross validated correlations for the three samples. All of the cross validated correlations were significant for the aggregate sample. These correlations ranged from a low of -.12 (p < .001) for the Conscientiousness biodata scored scale to a high of -.23 for the Total biodata scored scale. For the sheriff's group the correlations for the biodata scored scales for Neuroticism (r = -.19, p < .001), Openness to Experience (r = -.19, p < .001), Conscientiousness (r = -.17, p < .001), and Total (r = -.26, p < .001) were significant, and the scales for Agreeableness (r = -.12, p = .03) and Extraversion (r = -.13, p = .02) approached significance. The correlations for the probation department employees exhibited the least amount of shrinkage. These correlations ranged from a low of -.23 (p < .001) for the Agreeableness biodata scored scale to a high of -.38 (p < .001) for the Total scale.

The pairs of correlations for biodata and personality scoring were tested to determine if there were significant differences and these results are presented in Table 29. The empirical scales for

<table>
<thead>
<tr>
<th></th>
<th>Aggregate</th>
<th>Sheriff</th>
<th>Probation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>Z = 2.31, p = .01</td>
<td>Z = 3.26, p &lt; .001</td>
<td>Z = 2.26, p &lt; .001</td>
</tr>
<tr>
<td>C.I.</td>
<td>.04 - .22</td>
<td>.09 - .27</td>
<td>.05 - .32</td>
</tr>
<tr>
<td>Open/Exp</td>
<td>Z = 1.91, p &lt; .001</td>
<td>Z = 2.38, p &lt; .001</td>
<td>Z = 1.87, p &lt; .001</td>
</tr>
<tr>
<td>C.I.</td>
<td>.01 - .20</td>
<td>.04 - .20</td>
<td>.01 - .31</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Z = 3.33, p &lt; .001</td>
<td>Z = 2.60, p &lt; .001</td>
<td>Z = 2.31, p &lt; .001</td>
</tr>
<tr>
<td>C.I.</td>
<td>.09 - .27</td>
<td>.07 - .30</td>
<td>.07 - .43</td>
</tr>
<tr>
<td>Agreeable</td>
<td>Z = 3.14, p &lt; .001</td>
<td>Z = 1.56, p &lt; .001</td>
<td>Z = 2.20, p &lt; .001</td>
</tr>
<tr>
<td>C.I.</td>
<td>.09 - .29</td>
<td>-.01 - .21</td>
<td>.05 - .39</td>
</tr>
<tr>
<td>Consc</td>
<td>Z = 1.20, p &lt; .001</td>
<td>Z = 2.31, p &lt; .001</td>
<td>Z = 3.29, p &lt; .001</td>
</tr>
<tr>
<td>C.I.</td>
<td>-.01 - .08</td>
<td>.04 - .26</td>
<td>.11 - .34</td>
</tr>
</tbody>
</table>

Neuroticism (Z = 2.31, p = .01), Openness to Experience (Z = 3.33, p < .001), and Agreeableness (Z = 3.14, p < .001) were superior to the personality scales for the aggregate group. The Conscientiousness biodata scored scale (r = -.12) was not significantly better than the personality scale (r = .08; Z = 1.20, p = .11), and the Extraversion biodata scored scale (r = -.14) did not differ from the personality scale (r = .04) at the .01 level of significance (Z = 1.91, p = .03). For the sheriff's group all of biodata scoring correlations were significantly superior (Zs > 2.31, p < .001) to their personality counterparts with the
exception of Agreeableness ($Z = 1.56, p = .06$). For the probation group the biodata scored scale
correlations were significantly greater than the personality scale correlations for Neuroticism ($Z = 2.26,$
p = .01), Openness to Experience ($Z = 2.31, p = .01$), Agreeableness ($Z = 2.20, p = .01$),
Conscientiousness ($Z = 3.29, p < .01$), and the differences for Extraversion ($Z = 1.87, p = .03$)
approached significance. The results above demonstrated that the biodata scoring correlations for
Conscientiousness were better predictors than the personality correlations for two of the three groups
and therefore provide partial support for the hypothesis. Strong support was demonstrated for the
ability of biodata scoring to improve the validities of the Big Five scales.

As an additional analysis the Conscientiousness biodata scored scale was compared to the
Agreeableness, Openness to Experience, and Neuroticism biodata scored scales. For this a priori
comparison the former scale was assigned a weight of 1.0 and the other scales a weight of -.33. The
result of this contrast was not significant for the probation ($Z = .35, p = .86$), sheriff ($Z = .09, p = .96$),
or aggregate ($Z = -1.62, p = .052$) samples. Moreover, in the case of the latter comparison the
Conscientiousness biodata scored scale was actually inferior to the other scales.

A regression analysis was conducted to determine whether the biodata scored scales provided
incremental variance above and beyond that accounted for by the personality scales. In addition, a
supplementary analysis for each group was conducted whereby the ten scales were entered
simultaneously. The beta weights and the significance levels for the ten scales for the aggregate,
sheriff, and probation samples are presented in Table 30.

For the aggregate sample the Conscientiousness personality scale entered the equation in the first
stepwise block that produced a $R^2$ of .01 [$F(1, 553) = 3.96, p = .05$]. The biodata scored scales for
Openness to Experience and Neuroticism entered the equation in the next block and improved the $R^2$ to
.05 [$F(3, 551) = 10.03, p < .001$]. As a supplementary analysis the ten scales were entered as a
stepwise block yielding a $R^2$ of .07 [$F(10, 544) = 4.08, p < .001$]. The beta-weights for the Openness to
### Table 30. Supplementary Regression Analysis for CPI Biodata Scales and Turnover.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Aggregate (n = 555)</th>
<th>Sheriff (n = 308)</th>
<th>Probation (n = 134)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>Sig T</td>
<td>β</td>
</tr>
<tr>
<td>Neur</td>
<td>.00</td>
<td>.99</td>
<td>.11</td>
</tr>
<tr>
<td>Extra</td>
<td>-.07</td>
<td>.13</td>
<td>.08</td>
</tr>
<tr>
<td>Opn/Exp</td>
<td>.13</td>
<td>.01</td>
<td>.09</td>
</tr>
<tr>
<td>Agree</td>
<td>-.05</td>
<td>.46</td>
<td>-.06</td>
</tr>
<tr>
<td>Consc</td>
<td>-.03</td>
<td>.63</td>
<td>-.12</td>
</tr>
<tr>
<td>Bio-Neu</td>
<td>-.07</td>
<td>.16</td>
<td>-.20</td>
</tr>
<tr>
<td>Bio-Ext</td>
<td>-.04</td>
<td>.39</td>
<td>-.12</td>
</tr>
<tr>
<td>Bio-Opn</td>
<td>-.14</td>
<td>.01</td>
<td>-.14</td>
</tr>
<tr>
<td>Bio-Agr</td>
<td>-.09</td>
<td>.11</td>
<td>.01</td>
</tr>
<tr>
<td>Bio-Con</td>
<td>-.09</td>
<td>.58</td>
<td>-.13</td>
</tr>
</tbody>
</table>

Experience personality scale (β = .13, p = .01) and the Openness to Experience biodata scored scale (β = -.14, p = .01) were significant.

None of the personality scales entered the equation for the first stepwise block conducted for the sheriff group. The biodata scored scales for Openness to Experience and Neuroticism entered in the second stepwise that resulted in a significant $R^2$ of .05 [$F(2, 305) = 8.68$, $p < .001$]. When the ten scales were entered simultaneously the resulting $R^2$ of .10 was significant [$F(10, 297) = 3.26$, $p < .001$]. Significant beta weights were obtained for the Neuroticism biodata scored scale (β = -.20, p = .01) and the Conscientiousness biodata scored scale (β = -.13, p = .04).

For the probation group, none of the personality scales entered the first block. Entering the biodata scored scales increased the $R^2$ to .13 as a result of the Openness to Experience and Neuroticism scales [$F(2, 131) = 9.89$, $p < .001$]. For the follow-up analysis the set of ten scales was entered as a block. The beta-weights for the ten scales show that the Conscientiousness personality scale (β = -.32, p = .01), the Neuroticism biodata scored scale (β = -.24, p = .02), the Openness to Experience biodata...
scored scale ($\beta = -.20$, $p = .04$) and the Conscientiousness biodata scored scale ($\beta = -.30$, $p = .02$) were significant.

As a whole, the results above indicate that the biodata scored scales provide incremental variance over the personality scales for turnover. However, for the groups indicated above, results suggest that the Neuroticism and Openness to Experience biodata scored scales are better predictors of turnover than the Conscientiousness biodata scored scale.

Prototypical profiles were developed for those who left the organization and those who remained for the aggregate, sheriff, and probation groups. Hypothesis 10 specified that the PAMS profiles for those who remained in an organization would contain high elevations for Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability. The dimension values for the stayers and quitters for the three groups for the personality and biodata scored scales are presented in Table 31.

<table>
<thead>
<tr>
<th>Table 31. PAMS Profiles for Stayers and Quitters for Aggregate, Sheriff, and Probation Groups.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aggregate</strong></td>
</tr>
<tr>
<td><strong>Dim 1</strong></td>
</tr>
<tr>
<td><strong>Scale</strong></td>
</tr>
<tr>
<td>Neurot</td>
</tr>
<tr>
<td>Extrav</td>
</tr>
<tr>
<td>Openne</td>
</tr>
<tr>
<td>Agreea</td>
</tr>
<tr>
<td>Consct</td>
</tr>
<tr>
<td>Bio_N</td>
</tr>
<tr>
<td>Bio_O</td>
</tr>
<tr>
<td>Bio_A</td>
</tr>
<tr>
<td>Bio_C</td>
</tr>
</tbody>
</table>

Note: * Elevation greater than one standard deviation is depicted in the profile interpretation.

and are depicted in Figures 3a - 3d on page 120 for the aggregate sample, Figures 4a - 4d on page 121 for the sheriff sample, and Figures 5a - 5d on page 122 for the probation sample.
120

Figure 4a - 4d. PAMS Profiles for Personality and Biodata for Aggregate Sample.
Sheriff's Profile of Stayers

Figure 5a - 5d. PAMS Profiles for Personality and Biodata for the Sheriff's Department.
Probation's Profile of Stayers  

Probation's Profile of Terminaters  

Figure 6a - 6d. PAMS Profiles for Personality and Biodata for the Probation Department.
For the aggregate sample a two-dimensional model fit the data well for those who left the organization (quitters; STRESS = .06) as well as for those who remained in the organization (stayers: STRESS = .007). For dimension one, the profile for stayers (depicted in Figure 4a) and quitters (depicted in Figure 4b) was similar. Both profiles were driven by a high elevation for Neuroticism coupled with a negative elevation for Agreeableness. This may reflect a general personality profile indicative of law enforcement occupations defined as being well adjusted and not being able to easily get along with others. The second profile produced the most dramatic differences for the Conscientiousness scale. Here, the profile stayers had a positive elevation for the scale, whereas the quitters had a negative elevation for the scale.

A similar set of profiles was found for the sheriff and probation groups. In terms of model fit, a two-dimensional model was appropriate for stayers (STRESS = .004) and quitters (STRESS = .002) for the sheriff (depicted in Figures 5a and 5b, respectively), and stayers (STRESS = .005) and quitters (STRESS = .003) for probation (depicted in Figures 6a and 6b, respectively). For both the sheriff and probation the first profile dimensions were similar with a positive elevation for Neuroticism and a negative elevation for Agreeableness. The differences between the stayers and quitters occurred with dimension two as the stayers had a positive elevation for Conscientiousness and the quitters had a negative elevation for Conscientiousness.

Results from the three groups provide partial support for Hypothesis 10. Whereas Neuroticism and Agreeableness were important scales for stayers and quitters, the scales did not distinguish between the two. As predicted Conscientiousness produced the distinction between all stayers and quitters for the sheriff, probation, and aggregate samples. Also, as predicted Extraversion had an insignificant role in the profiles of both stayers and quitters.

Hypothesis 11 proposed that the biodata scoring elevations would be more pronounced for Conscientiousness, Agreeableness, Neuroticism, and Openness to Experience. This analysis was also conducted separately for the aggregate, sheriff, and probation samples.
For the aggregate sample a two-dimensional model was appropriate for the biodata scoring profiles for the stayers (STRESS = .006) and quitters (STRESS = .07). These profiles are shown in Figures 4c and 4d on page 120. Unlike the personality dimensions, the stayer and quitter biodata scoring profiles had differences for both dimensions. The first dimension for stayers had a positive elevation for the biodata scored Neuroticism scale, whereas the profile dimension for quitters had a positive elevation for Openness to Experience biodata scored scale and a negative elevation for the Extraversion biodata scored scale. The second dimension for stayers had a positive elevation for Openness to Experience, whereas the quitters' profile was driven by a negative elevation for the Neuroticism biodata scored scale.

A two-dimensional model also provided a good fit for the sheriff sample as indicated by the STRESS statistic of .004 for the stayers and .002 for the quitters. These profiles are depicted in Figures 5c and 5d on page 121. Dimension one for the stayer profile had a positive elevation for Neuroticism and a negative elevation for Openness to Experience, whereas the quitter profile had a negative elevation for Neuroticism and a positive elevation for Openness to Experience. Dimension two also produced differences with the stayer profile having a positive elevation for Conscientiousness and the quitter profile having a positive elevation for Neuroticism.

Finally, the probation sample also conformed to the two-dimensional interpretation for the stayer (STRESS = .004) and quitter (STRESS = .005). These profiles are shown in Figures 6c and 6d, respectively. The dimension one profile for stayers had a positive elevation for the Extraversion biodata scored scale and a negative elevation for the Openness to Experience biodata scored scale. The dimension two profile for quitters also had a positive elevation for the Extraversion biodata scored scale but instead had a negative elevation for Agreeableness. For the second dimension the stayer profile was defined by a positive biodata scored Neuroticism elevation and a negative biodata scored Conscientiousness elevation and the quitter profile was defined by a negative biodata scored Openness to Experience elevation.
Although the biodata scoring profile dimensions differed for the stayer and quitter groups, the elevations for the Neuroticism, Agreeableness, Openness to Experience, and Conscientiousness biodata scoring profiles were not more pronounced than similar elevations produced for the personality scales.

Hypothesis 12 specified that the level parameter for the biodata scored personality measure would explain more variance in the turnover criterion than the level parameter for the traditionally scored scale. The personality level parameter ($X = .03, \text{s.d.} = .45$) for the aggregate sample produced a $R^2$ of .00 that was not significant [$F(1, 553) = .28, p = .59$]. Entering the biodata scoring level parameter ($X = -.02, \text{s.d.} = .70$) in the second block improved the $R^2$ to .05 [$F(2, 552) = 14.98, p < .001$].

For the sheriff sample, the $R^2$ (.00) for the personality level parameter ($X = .01, \text{s.d.} = .46$) was not significant [$F(1, 306) = .00, p = .98$], but the $R^2$ (.06) for the biodata scoring level parameter ($X = -.01, \text{s.d.} = .65$) was significant [$F(2, 305) = 9.71, p < .001$]. Finally, for the probation sample the personality level parameter ($X = -.02, \text{s.d.} = .45$) produced a $R^2$ of .00 that was not significant [$F(1, 132) = .17, p = .68$], but entering the biodata scoring level parameter ($X = -.03, \text{s.d.} = .63$) improved the $R^2$ to .16 that was significant [$F(2, 131) = 12.88, p < .001$]. These findings suggest that the overall measure of biodata scoring explains more criterion variance in the turnover variable than the overall personality parameter.

**Summary**

Results from Sample Three provided mixed support for the hypotheses. First, the personality scale for Conscientiousness did not reliably predict absenteeism across the six years that data were collected. Second, as predicted, biodata scoring improved prediction of absenteeism for all six years. At least four of the biodata scored scales were significant predictors for each year that data were collected. Follow-up analyses demonstrated at least two of the biodata scored scales were significantly better at predicting absenteeism than the personality scales for each year. The best results were obtained in 1996 when all of the biodata scored scales outperformed the personality scales. In addition, the results for the stepwise regression analyses demonstrated that biodata scoring explained variance
incrementally to the personality scales; although the same biodata scored scales did not always enter the equation.

The personality profiles for the low absenteeism group conformed to a two-dimensional configuration. For all six years, the first dimension contained an elevation for Neuroticism. An elevation for Agreeableness was present for five years. Similarly, dimension two contained an elevation for Conscientiousness for five of the six years. Although not conclusive, the results do suggest that Neuroticism, Agreeableness, and Conscientiousness are important components of the low absenteeism profile.

The biodata scoring profile dimensions did not contain more pronounced elevations as predicted by the hypothesis. For five of the six years, the Openness to Experience component was important for the first dimension. Neuroticism and Agreeableness were present for three years and Conscientiousness was only important for the first year of analysis. For dimension two, Neuroticism was present in the low absenteeism profile for four of the six years, Extraversion for three of the six, and Conscientiousness for two of the six. Given that biodata scoring changes the scoring of the scales and possibly their interpretation, it is difficult to determine if the scales measure the same initial construct without additional construct validation studies.

Further support for the argument that measurement has changed the scales is evident when comparing the level parameters for the personality and biodata scored scales. In every case, the biodata scoring level parameter explained variance in the absenteeism criterion beyond the personality level parameter. With the exception of 1992, both the high and low absenteeism individuals conformed to the profiles equally well as exhibited by the examination of the two group’s mean correspondence weights. Mixed support was obtained for the importance of the biodata scored Conscientiousness component that outperformed the biodata scored Neuroticism and Agreeableness component for three of the six years. One possible explanation, supported by the repeated measures ANOVA, is that the items and the weights that contributed to the scale differed as a function of years. Therefore, the inclusion or weighting of a particular item was contingent upon the samples used to derive the scales.
The weighting of items was not biased in favor of the Conscientiousness component as the mean of the weights was similar across the six years with the exception of 1992.

For turnover, none of the personality constructs were significant predictors for the aggregate, sheriff, or probation samples. Biodata scoring significantly improved the predictive ability as the biodata scored scales for Neuroticism, Openness to Experience, and Agreeableness predicted absenteeism for the Aggregate sample; Neuroticism, Openness to Experience, Agreeableness, and Conscientiousness predicted absenteeism for the Probation sample; and all five biodata scored scales predicted absenteeism for the Sheriff sample. Biodata scored scales consistently provided incremental variance above that explained by the personality scales. Specifically, the biodata scored scales for Openness to Experience and Neuroticism entered in the second block of the stepwise. The only personality scale to enter in the first block was the Conscientiousness scale for the aggregate sample. As noted, the aggregate sample is comprised of a cross-section of occupations and not just the law enforcement classes found in the sheriff and probation groups. Therefore, one possible explanation of why the Conscientiousness was significant for this sample but not the others may be that this component of personality is more important for clerical or administrative functions rather than law enforcement functions (e.g., Day & Silverman, 1989).

The personality profiles for the individuals who remained in the organization contained one dimension with a positive elevation for Neuroticism and a negative elevation for Agreeableness. A second dimension for stayers contained a positive elevation for Conscientiousness. These profiles were replicated for the three samples and therefore provide support for the hypothesis.

The biodata scoring profiles did not contain more pronounced elevations for the specified components. Instead, the biodata scoring profiles differed for the three samples. The aggregate sample had one dimension driven by Neuroticism and a second by Openness to Experience; the sheriff sample had one dimension defined by a combination of Neuroticism and Openness to experience and a second dimension containing Conscientiousness; and the probation sample had a positive elevation for Extraversion coupled with a negative elevation for Openness to Experience for the first dimension, and
a positive elevation for Neuroticism and a negative elevation for Conscientiousness for the second dimension.

To compare the variance explained by the personality and biodata scoring profiles the level parameters of each were compared. For all three samples the personality level parameter did not significantly predict turnover. By comparison, the biodata scoring level parameter predicted turnover for all three samples.

Taken as a whole, the findings reported above demonstrated the effectiveness of biodata scoring to predict absenteeism and turnover. These results suggest that biodata scoring is influenced by situational characteristics, such as the sample used to derive the scales and the specific occupational group used, that may impact the composition of the scales and the subsequent validity coefficient. Biodata scoring may also change the constructs being measured as shown by the incremental validity of the scales and the revised structure of PAMS profiles caused by changes in scoring.
GENERAL DISCUSSION

In this section I integrate results from the three samples. First, results for the personality correlations and profile analyses for absenteeism and turnover are discussed. Second, biodata correlational results and profile analyses are presented. Third, the limitations of the study are summarized followed by suggestions for future studies. Finally, practical implications are presented.

PERSONALITY FINDINGS

Personality-Absenteeism Correlations

Results supporting the relationship between personality and absenteeism were mixed. The results for the first two samples did not support the absenteeism-personality relationship. The magnitude of the correlations were not only small, they were also not consistent for the development and hold-out groups. As such, Samples One and Two provided disconfirming evidence to the literature that has demonstrated the importance of Conscientiousness, Agreeableness, and Emotional Stability for predicting counterproductive behavior, including absenteeism (Barrick & Mount, 1991; Tett et al. 1991).

Guarded support for the hypothesis was found for Sample Three. Conscientiousness was associated with absenteeism for the hold-out and development samples in all but two of twelve instances and the average $r$ taken across all years was 13.3. These correlations, though small (Cohen & Cohen, 1988), are consistent with Barrick and Mount’s (1991) meta-analysis. They found that the uncorrected mean $r$ for Conscientiousness and personnel data (i.e., absenteeism) taken across occupations was .13 and that within the occupational group labeled “Police” the uncorrected mean $r$ was also .13.

None of the other specified personality dimensions were significantly associated with absenteeism for Sample Three. The means for the uncorrected correlations of the twelve samples from this study were .07 for Extraversion, .02 for Emotional Stability, -.01 for Agreeableness, and -.01 for Openness to Experience. By comparison, the average mean observed correlation coefficients reported by Barrick...
and Mount were .05, .05, .04, and .03 for Extraversion, Emotional Stability, Agreeableness, and Openness to Experience, respectively for personnel data (e.g., absenteeism).

Given Barrick and Mount’s (1991) standards, the results for Sample Three would appear to support the use of Conscientiousness to predict counterproductive behavior. However, one might question the practical utility of a predictor that explained only 1.7% of the variance in the criterion. It must be noted that the correlations reported in the present study were not corrected for criterion or predictor unreliability that might have increased the proportion of variance explained. Corrections were not made in this study based upon the recommendation of Hogan, Hogan, and Roberts (1996) who stated: “In particular, we believe that in some cases, corrections for attenuation are used inappropriately and lead to an overestimate of predictor-criterion relationships” (p 471).

In sum, an association between personality and absenteeism was not supported for Samples One and Two, and the correlations found for Sample Three were small (Cohen & Cohen, 1988) and were disconfirmed in two of the 12 samples. As such, the present study offers minimal support for an association between Conscientiousness and absenteeism (Barrick & Mount, 1996; Ones, Viswesvaran, & Reiss, 1996) and no support for the other hypothesized personality variables.

Personality-Absenteeism Profiles

The expectation that the personality profiles of low absenteeism individuals would be characterized by high Conscientiousness, high Emotional Stability, and high Agreeableness elevations was not supported for any of the three samples. In addition, whereas the personality profile for the high and low absenteeism group diverged, the divergence was inconsistent for the three samples. For example, the low absenteeism profile for Sample One was characterized by low elevations for Ambition and Likeability and high elevations for Intellectance and Prudence. Consequently, an individual with these characteristics may not engage in absenteeism behavior due to a high work ethic and conformist attitude as noted by the high Prudence score, a lack of ambition that results in either enjoyment of the work of school bus driver or a lack of interest in other job alternatives as indicated by the low
Ambition score, a tough interpersonal style that may be a requirement of an inner city school bus driver reflected by the low Likeability score, and a willingness to accept novel experiences characterized by the high Intellectance score. A second low absenteeism profile also contained a high elevation for Intellectance combined with a negative elevation for Prudence. This profile is defined as being imaginative, inventive, and quick witted, but easily bored and inattentive to detail; and being undependable, disorganized, careless, impulsive, but flexible and accepting of change. One possible explanation to account for this profile is that an employee may exhibit a low absenteeism rate because high absenteeism may result in the employee's termination due to the low score on the Prudence dimension.

On the other hand, one high absenteeism profile consisted of individuals who are low Adjustment, high Prudence, and high Sociability. These individuals are tense, moody, temperamental, unhappy, and easily stressed; are outgoing, colorful, gregarious, and enjoy social attention; and are dependable, organized, easy to supervise, but somewhat inflexible and resistant to change (Hogan & Hogan, 1992). This profile dimension may reflect one who is absent for legitimate reasons as indicated by the low adjustment scores; or the conscientious person who is absent because demands imposed by family obligations or health are deemed more important than work; or is absent because not exposing others to illness is perceived as the prudent thing to do. This profile may reflect either (a) someone who is genuinely sick or (b) someone who is willing and able to come to work but is absent because of other legitimate reasons. The second profile dimension contained high Likeability, high Sociability, and low Intellectance elevations. This pattern is defined by those who are outgoing, colorful, and gregarious; are friendly, warm, and popular; and are unimaginative, narrow, and tolerant of boredom (Hogan & Hogan, 1992). According to Hogan and Hogan, persons with high scores on Likeability and Sociability are delightful, entertaining, and like parties. It is possible that this profile accounts for individuals who are absent from work to either attend social functions, or recover from them. Future
studies need to assess the reason for absenteeism to shed light on why particular profiles exhibit high levels of absences.

These results suggest that both high and low Prudence individuals may engage in absenteeism behavior. Because of the many reasons for absenteeism (Nicholson & Payne, 1987), it is possible that the highly prudent individual engages in specific types of absence behavior more frequently than someone with lower prudent scores. For example, if one were to make a distinction between voluntary and involuntary absenteeism (Hackett & Guion, 1985), a low prudent individual may be more likely to miss work as a result of voluntary absenteeism, such as taking a Friday off to extend the weekend. On the other hand, a high prudent individual may miss work due to an actual illness or illness of a family member because of involuntary absenteeism. However, it is possible that both individuals may exhibit the same rate of absenteeism.

The profiles for the low and high absenteeism groups also diverged for Sample Two. Here, one dimension of the profile for low absenteeism had high Adjustment, high Likeability, and low Prudence elevations. These individuals are calm, self-confident, upbeat, even-tempered and handle pressure well; are friendly, warm, and popular; but have the negative traits of being undependable, disorganized, careless, and impulsive (Hogan & Hogan, 1992). The second dimension of the profile for the low absenteeism group had no marked elevations and was not interpreted. It is interesting to note that the first dimension for the high absenteeism group contained the same Adjustment, Likeability, and Prudence elevations as the low absenteeism group. Consequently, it may be that the well adjusted, well liked individual, who does not have a dependable nature, but is willing to accept change, may gravitate to transit bus driver occupations. However, it appears that this dimension of the profile is not associated with absenteeism. Therefore, one might argue that because this dimension is present for both the high and low absenteeism groups it may be indicative of a general occupational profile for transit bus drivers.
The second dimension of the profile for the high absenteeism group had a low Intellectance, high Prudence pattern of elevations that may be interpreted two ways. One interpretation, based on ASA literature would suggest that individuals who do not conform to dimension one of the high and low absenteeism groups are not suited to the occupation and hence use absenteeism as a withdrawal mechanism (Staw & Oldham, 1978). A second interpretation would be that high Prudence individuals use absenteeism because they view other obligations, such as personal or family health, as more important than work.

Sample Three compared the high absenteeism profile to the low absenteeism profile for six years. For the low absenteeism group, the pattern for the first profile dimension was replicated for all six years. This dimension conformed to a high Neuroticism and low Agreeableness pattern of scores and represents individuals who are calm, relaxed, even-tempered, and are able to face stressful situations without becoming upset or rattled; and tend to be egocentric, skeptical of others’ intentions, and competitive rather than cooperative (Costa & McCrae, 1992). The first dimension of the high absenteeism group also contained high Neuroticism and low Agreeableness elevations with the exception of the 1992 sample. In 1992, the dimension did contain a low Agreeableness elevation, but it also contained a high Extraversion elevation. One possible explanation is that the Extraversion elevation may have offset the high Neuroticism factor due to the ipsative nature of PAMS (Davison, 1996). Unfortunately, because PAMS is exploratory (Davison et al., 1996), a confirmatory approach could not be employed to determine whether the high Neuroticism, low Agreeableness model would have been appropriate for the data. Another explanation is that the dimension may explain actual differences between the low and high absenteeism groups. Extraverted individuals who are outgoing, disagreeable, and antagonistic may be less suited to the work of law enforcement and hence use absenteeism as a withdrawal characteristics due to poor person-job fit. Alternatively, someone who is outgoing and antagonistic may miss work as a result of being involved in altercations or disputes.
The interpretation of the second dimension for the high and low absenteeism profiles is more difficult. For the low absenteeism group, the only meaningful elevation for all six years occurred for the Conscientiousness factor. In three years it was a positive elevation (1993, 1995, and 1996) and in three years it was a negative elevation (1991, 1992, and 1994). Davison (1996) has indicated that it is sometimes necessary to interpret the mirror image of the profile due to the ipsative scaling properties of PAMS. Using Davison's suggestion, if one were to rotate the profiles for 1991, 1992, and 1994 a similar pattern of high Conscientiousness scores would be obtained and the profiles would be consistent. This rotation appears logical in light of the finding that high scores on Conscientiousness were associated with low scores on absenteeism; however, future studies are needed to determine the validity of this contention.

The second dimension of the high absenteeism profile did not contain consistent marked elevations. A low Openness to Experience elevation was reported in 1991 and 1993, a low Neuroticism, high Openness to Experience combination occurred in 1992, and a high Conscientiousness elevation was found in 1995. Therefore, one cannot conclude with confidence that any of the personality scales reliably differentiated between low and high absenteeism group profiles.

In summary, both the high and low absenteeism groups appeared to have similar high Neuroticism (well adjusted), low Agreeableness profiles that may have been a result of the psychological screening required of employees entering law enforcement occupations. As such, the only dimensions allowing for variability were the Openness to Experience, Conscientiousness, and Extraversion factors, but these factors did not consistently explain differences between the two groups’ profiles. The results of the low absenteeism profiles therefore did not support for the hypothesis. The Conscientiousness scale elevations were unreliable and varied from sample to sample. In addition, the results for the high absenteeism group were equally inconsistent and future studies are needed to determine the usefulness of PAMS methodology using other samples, occupations, and organizations.
Taken together, the results for the three samples suggest that the profiles for high and low absenteeism groups did differ, but not as expected. In addition, Sample Three results were inconsistent and future studies are warranted to assess the usefulness of using profile analysis to explain absenteeism.

**Personalitv-Absenteeism Profile Correspondence**

Support for the hypothesis that the low absenteeism group would more closely match the predominant profile than the high absenteeism group was not supported. For all three samples, the high and low absenteeism group's degree of correspondence to the predominant profile did not differ. According to Davison et al. (1996) this suggests that the group's correspondence to the profile does not exhibit an association with absenteeism; although other profile characteristics such as the level parameter and the scatter and shape of the profile may be correlates.

**Personalitv-Turnover Correlations**

The results for the association between personality and turnover were contrary to expectation. For Sample One, the findings did not support the hypothesized correlation between personality and turnover and therefore did not support the results obtained by Barrick and Mount (1996) or Hough et al. (1990). The former reported personality-turnover correlations of -.23 for Conscientiousness and -.20 for Emotional Stability, and the latter .24 for Achievement and .17 for Dependability for job involvement. For Sample One, with the exception of Adjustment (r = .15) for the hold-out sample, none of the scales produced correlations above .07.

Sample Three results also did not provide support for an association between personality and turnover. Using an aggregate sample, the Conscientiousness scale exhibited the largest correlation between personality and turnover (r = .08). Some significant correlations did exist for either the hold-out or development samples, but these relationships were not found for both groups. For example, Neuroticism and Extraversion were significantly associated with probation department turnover for the development group; and Openness to Experience, Extraversion, and
Conscientiousness were significantly associated with sheriff's department turnover for the development group. One would expect the measures to be equally predictive of turnover for both development and hold-out groups which was not the case here. Hence, unequivocal support for the correlation between personality and turnover was not demonstrated in these samples.

**Personality-Turnover Profiles**

For Samples One and Three, there were differences noted between the stayer and leaver groups. However, as with the absenteeism measure, the profiles did not contain the expected elevations for Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability.

For the school bus drivers in Sample One, the PAMS profiles for employees who remained in the organization conformed to either of two dimensions. The first dimension contained low Ambition, low Likeability, and High Intellectance elevations. Hogan and Hogan (1992) defined these individuals as being quiet, unassertive, not energetic, and not interested in advancement; unresponsive, critical, and tough; and imaginative, inventive, and quick witted, but easily bored and inattentive to detail. Following an ASA interpretation, this may represent the type of individual that not only gravitates to an occupation but does not leave.

The second profile dimension for the stayer contained high Adjustment, low Prudence, and low Sociability elevations. Hogan and Hogan (1992) ascribe the following characteristics to these three personality factors: calm, self-confident, upbeat, even tempered, and handle pressure well; undependable, disorganized, careless, impulsive, but flexible and accepting of change; and shy, withdrawn, aloof, and keeps others at a distance. These individuals may remain in the organization because (a) they are able to handle the stress of being an inner city school bus driver, (b) their low scores on the Prudence dimension make them unattractive to other employers, and (c) the environment of dealing with children rather than adults fits well with their low Sociability scores.

By contrast, profile one for those that left the organization had high Likeability, high Sociability, low Intellectance, and low Prudence elevations. These are individuals who are friendly, warm and
popular, are tolerant of boredom, and are undependable and disorganized. Because these qualities were also noted in the second dimension of the profile for the high absenteeism individual, with the exception of low Prudence, it is possible that these individuals left the organization through involuntary termination. Specifically, the combination of a low work ethic with the characteristics of a high absenteeism profile may lead to termination; whereas, the combination of a low work ethic combined with the characteristics of low absenteeism profile (e.g., high Intellectance) may not.

A second profile dimension for those who left the organization had low Intellectance, high Prudence elevations. According to Hogan and Hogan (1992), low Intellectance individuals thrive in entry-level and mid-level jobs that require close attention to detail and the ability to tolerate boredom, citing examples such as air traffic controller, nuclear power plant operator, and locomotive engineer. It may have been the case that individuals were attracted to this position because of an expectation that the work would be relatively routine and boring, only to find out that the demands of an inner school bus driver include frequent parental and student confrontations (Lin et al., 1990). Because their high Prudence characteristics may have been valued by other organizations, the poor person-job fit may have resulted in voluntary turnover for this profile.

Sample Three was also used to investigate the relationship between personality profiles and turnover. The results provided support for the role of Conscientiousness in turnover. The aggregate, sheriff, and probation groups all contained a similar profile one dimension for both stayers and leavers. Similar to the absenteeism profile reported earlier, it is possible that the first profile dimension reflected a personality type necessary to fulfill the occupational requirements. This dimension of the profile contained a high Neuroticism, low Agreeableness pattern that has been shown by others to be indicative of success in law enforcement (e.g., Lorr & Strack, 1994; Mills & Bohannon, 1980). Where the stayer and leaver groups diverged most dramatically was with the second profile dimension. Here the stayer groups for all three samples had a positive elevation for Conscientiousness, whereas the three quitter groups had a negative elevation for Conscientiousness.
As such, support for the role of Conscientiousness for turnover was inconclusive. For Sample One, Conscientiousness was not a characteristic difference between the stayer and leaver groups. One possible explanation for these differences is that a distinction was not made between voluntary and involuntary turnover. Campion (1991) has stated that each type of turnover is likely to have different antecedent variables and this may explain why the profiles did not contain the expected elevations. Only voluntary turnover was assessed for Sample Three and these results suggest that Conscientiousness may explain differences between the profiles of stayers and leavers.

**Personality-Turnover Profile Correspondence**

For Sample One and Three, no differences were noted between the stayer and leaver group’s correspondence weights. Therefore, although the profiles diverged in terms of the scatter and shape (Davison et al., 1996), stayers and leavers did not differ in the manner in which they corresponded to the profile. It was expected that the stayer group would correspond more closely to the profile based on the theory that organizations become increasingly homogenous over time (Jordan et al., 1991; Schneider, 1987), but this was not the case. In sum, for Sample Three there appeared to be one dimension of a profile that is characteristic of law enforcement personnel and the main characteristic of the second dimension of the profile that differentiates the attrition group from the stayer group is Conscientiousness.

**Personality Summary**

Given the results summarized above, it is difficult to make hard and steadfast conclusions regarding the association of personality with either absenteeism or turnover. In some cases the correlations were not significant, in others the correlations were significant for one sample and not another, and in some cases the correlations were inconsistent from year to year.

A similar conclusion could be made for the personality profile analyses. A prototypical profile for low absenteeism individuals containing Conscientiousness, Agreeableness, and Emotional Stability was not found for any of the three samples. Further, for turnover, a prototypical profile containing
Conscientiousness, Agreeableness, Emotional Stability, and Openness to Experience was not found for either Sample One or Three. The only consistent finding that offered partial support for the hypothesis was the turnover profiles for the sheriff, probation, and aggregate groups of Sample Three. For these samples it was shown that the Conscientiousness factor differentiated between stayers and leavers.

Combined, the results from the correlation and profile analyses did not support the contention that broad measures of personality, such as the Big Five, can predict broad measures of performance (Hogan et al., 1996), such as absenteeism or turnover. Instead, the results tend to support (a) the situational specificity argument advocated by Day and Silverman (1989) to carefully select both predictor and criterion measures for specific organizations and (b) the need to consider Block's advice about the Big Five (FFA) to undertake a "wider reflection" on the nature of personality and the nature of the FFA before embracing the latter" (p. 229).

**BIODATA FINDINGS**

**Biodata-Absenteeism Correlations**

The primary purpose of the study was to demonstrate that empirical scoring could improve the predictive validity of extant personality measures. It was expected that the cross validities for the biodata scored scales for Conscientiousness, Agreeableness, and Emotional Stability would be significantly associated with absenteeism. For Sample One, empirical weighting was able to produce significant correlations with absenteeism for the development group. Unfortunately, the cross validities were subject to shrinkage and the correlations for Adjustment ($r = -.12$), Prudence ($r = -.12$), Sociability ($r = -.17$), and Total ($r = -.18$) were not statistically significant for the hold-out group and at best explained 3.2% of the variance in absenteeism.

Sample Two results were similar. The developmental group correlations were significant, but excessive shrinkage of the cross validated scales resulted in the correlations not being significant. The
largest cross validities were obtained for Adjustment (r = -.10), Likeability (r = -.15), Sociability (r = -.12), and Total (r = -.17) explaining 1% to 2.9% of the variance in absenteeism.

Finally, all of the developmental correlations were significant for Sample Three. Although shrinkage occurred for the scales when cross validated for the hold-out group, most of the biodata scored scales remained significant. The average r for the six years of the sample were: Neuroticism (r = .12), Extraversion (r = .18), Openness to Experience (r = .21), Agreeableness (r = .15), Conscientiousness (r = .20), and Total (r = .28).

**Incremental Validity of Biodata Absenteeism Scales**

For Sample One, the biodata scored scales provided incremental validity over the personality scales; however, Conscientiousness was not the personality or biodata scored scale that entered the equation. Incremental validity was not demonstrated for Sample Two as neither the personality scales nor the biodata scored scales entered the equation. For Sample Three, biodata scored scales did explain incremental validity above and beyond the personality scales. However, there was an inconsistency in the scales that entered the equation as shown in Table 32, and the variance explained ranged between 1% and 3%.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality</td>
<td>Neur. .02</td>
<td>N/A</td>
<td>Cons. .03</td>
<td>N/A</td>
<td>Cons. .02</td>
<td>Cons. .01</td>
</tr>
<tr>
<td>Biodata</td>
<td>Open .09</td>
<td>Open .04</td>
<td>Open .09</td>
<td>Open .06</td>
<td>Cons. .07</td>
<td>Cons. .10</td>
</tr>
</tbody>
</table>

for the personality scales and 4% and 9% for the biodata scored scales. The inconsistency of entry by the biodata scored scales is most likely a function of the empirical weighting of items. As noted, the weights and inclusion of items changed year-to-year, thereby influencing the content of the scales.
Biodata-Absenteeism Profiles

Both the high and low absenteeism groups conformed equally well to the predominant profiles and the groups did not differ in the profiles' ability to represent both groups (Davison, 1994, 1996). However, interpreting the biodata scoring profiles was somewhat problematic. Whereas the item content had remained the same, the inclusion and scoring of the items had not. Therefore, the assumption that biodata scored scales were equivalent measures of Big Five constructs was tenuous. Additional evidence for this argument can be found in the different patterns and elevations of scales that were produced using biodata scoring compared to those obtained for personality scoring. Moreover, there was variability within the biodata profiles as evidenced by the divergent profiles developed for the six years of absenteeism data. Because the first dimensions of the personality profiles were relatively consistent for the low absenteeism group during this same six year period, it is likely that the biodata scoring profile inconsistencies were due to the differential weighting and exclusion of items.

Biodata-Turnover Correlations

Overall, results were mixed for the hypothesis that biodata scored scales developed from extant personality measures would be significantly associated with turnover. For the development group in Sample One, scales were developed that were significantly correlated with turnover. However, the cross validities were not significant and the observed correlations of -.10 for Adjustment, -.12 for Sociability, -.15 for Likeability, and -.17 for Total explained from 1% to 2.9% of the variance in the criterion. As previously mentioned, no distinction was made between voluntary and involuntary turnover. It is conceivable that the members of these categories might have responded differentially to the PEPI items used in this study because of the different correlates of voluntary and involuntary turnover (Campion, 1991). As noted by Mumford and Owens (1987), empirical scales are typically more effective when there is clear and specific definition of the criterion measure.
Sample Three focused exclusively on voluntary turnover and used two separate occupational groups to assess the utility of developing scales for a specific occupational group (i.e., department). Of the three occupational groups studied, the probation department yielded the highest cross-validities and all of the biodata scored scales were significantly associated with turnover. Further, the correlation between the biodata scored Total scale and turnover ($r = -.38$) was the largest coefficient obtained in this study. All of the correlations between the biodata scored scales developed for the sheriff’s department and turnover were significant and the largest correlation was obtained for the Total biodata scored scale and turnover ($r = -.26$). Finally, the scales developed for the aggregate sample produced correlations that were significant, but were also consistently lower than the probation department correlations.

These findings provide a partial answer to the question posed by Rothstein et al. (1990) who developed scales across a broad range of organizations to predict job performance but wondered whether there was some loss in precision by not keying the scales for each specific organization. The results from the present study suggest that it may be better to develop scales within an organization, and if possible within a department to maximize prediction. Because this finding was obtained in one organization, and only two occupational groups, future investigations should attempt to specify the effects of developing scales designed for multiple organizations or multiple occupations compared to scales developed specifically for a situation.

**Biodata-Turnover Profiles**

Similar to the absenteeism criterion, the biodata turnover profiles differed from the personality turnover profiles. Biodata profile differences between the stayer and leaver groups for Sample One were most dramatic for the Intellectance and Prudence scales for dimension one and the Intellectance scales on dimension two. Although these scales resulted in differences in the profiles, they were not among the scales that exhibited the highest correlations with the criterion as shown in the correlational analysis.
The biodata profiles for the stayer and leaver groups of Sample Three also differed from the personality profiles. These differences were noted for the sheriff, probation and aggregate sample. In addition, the biodata profile dimensions differed between the occupational groups even though the dimensions for the personality profiles were similar. Whereas there was a lack of consistency for the biodata profiles, the personality profiles for the three groups were consistent. Combined, these results suggest further that biodata scales are measuring constructs that are not equivalent to those measured by personality scales.

Superiority of Biodata Scales

This study attempted to demonstrate that empirical scoring could improve the validity of personality measures. The findings for Sample One (absenteeism and turnover) and Sample Two (absenteeism) did not demonstrate the superiority of the biodata scales.

Results from Sample Three provided mixed support for the hypothesis that biodata scoring would produce a stronger association with absenteeism and turnover than traditional scoring. For Sample Three, the Conscientiousness biodata scale exhibited a larger correlation with the criterion than the personality scale for three of the six years of absenteeism data and was more effective for the sheriff and probation samples for the turnover measure. As such, the results are inconclusive and warrant further study.

Biodata scoring was also hypothesized to improve the scoring of the Neuroticism personality scale. Here, biodata scoring was superior for only one year of the six years of absenteeism data, but was superior for all three turnover groups. The Agreeableness biodata scale was hypothesized to outperform its personality scale counterpart. Here, the biodata scale was a superior predictor of absenteeism for three of the six years, and was a more effective predictor of turnover for the sheriff and aggregate sample. Finally Openness to Experience was hypothesized to be a more effective predictor of turnover and this was supported for the three turnover groups. Although these findings are
somewhat inconsistent, they are not unexpected as the weights and inclusion of items varied from year to year and from sample to sample.

Superiority of Conscientiousness Scales

Because of the previous findings that Conscientiousness was an important predictor of counterproductive behavior, it was expected that many of the items used in the empirical scales would come from this dimension and that these items would be assigned larger empirical weights. This hypothesis was not supported, with one exception, in this study. Consequently, one might conclude the propensity to turnover or to engage in absenteeism is comprised of a combination of personality factors that are not discernable at the Big Five level of analysis (Costa & McCrae, 1995). Therefore, personality relationships with criteria might best be studied at a lower level of analysis, such as the facet level (Costa & McCrae, 1995).

STUDY LIMITATIONS

Limitations of the present studies that may have impacted the results include the following: First, it is difficult to specify the effects that the volunteer incumbent participants of Sample One and Two had on the results compared to the applicant participants of Sample Three. It is purely speculation that the significant findings for Sample Three were attributable to applicant samples, whereas the findings that were not significant in Samples One and Two were due to incumbent samples.

Previous research suggests that these sample differences may be a cause for concern. For example, Schmit and Ryan (1993) and Cellar et al. (1996) showed that applicant and non-applicant responses differed for personality measures. Both sets of authors noted a sixth factor in applicant samples that was not present in incumbent groups. Moreover, Cellar et al. (1996) noted that the sixth factor in their sample did not contain the same Conscientiousness items found in the Schmit and Ryan study. Therefore, the structure of personality measures may differ from occupation to occupation as Cellar et al. concluded that differences with the sixth factor may have been a result of the composition of the two samples. Specifically, Schmit and Ryan's sample consisted of applicants receiving employment
assistance for the U.S. Employment Service, whereas Cellar et al.'s sample consisted of flight attendant trainees.

Further evidence for differences between applicant and incumbents can be found in the work of Schmit et al. (1993). They showed that when personality measures were responded to using an "at work" frame of reference the criterion validity increased. In a similar vein, Stokes et al. (1993) noted that there was a lack of item overlap between biodata keys developed for incumbent samples compared to applicant samples. They found that 20% or fewer of the biodata items developed for the incumbent sample were valid for the applicant sample. In addition, Stokes et al. showed that a measure of social desirability had a correlation of .36 with turnover for the applicant sample (n = 2,276) and a correlation of .23 for the incumbent sample (n = 2,262). Stokes et al. called into question the common practice of developing biodata keys on incumbents and assuming their generalizability to applicant samples.

The results from these studies suggest that applicant and incumbent sample responses to personality measures may differ. In addition, keys developed through empirical scoring may differ for incumbents and applicants. Consequently, these findings may explain the insignificant results for the incumbent participants of Sample One and Two, compared to the significant findings for the applicant participants of Sample Three. This is especially true for the empirically scored keys for Sample Three. Alternatively, the divergent findings may have been attributable to the occupations used, as different occupational groups may respond differently to personality measures (e.g., Cellar et al., 1996). The findings from the present study coupled with the results from Cellar et al. (1996), Schmit and Ryan (1993), Schmit et al. (1993), and Stokes et al. (1993) suggest that future research is needed to assess the potential impact of applicant versus incumbent samples for both the structure of personality tests and the relationship of personality and biodata scales with criteria.

A second limitation of the study that did not allow one to compare results from the three samples had to do with the personality measures used in this study. The personality measure used in Sample
One (PEPI) was still in the developmental stages. Many of the items were trial items and were not retained in future revisions. The personality measure used in Sample Two, the HPI, uses six factors to explain the Big Five personality structure. For Sample Three, the measure of the Big Five was the CPI. Items were not developed specifically to measure Big Five categories but were subsequently assigned to Big Five categories by McCrae et al. (1993) and therefore the CPI may have limitations as a measure of Big Five dimensions (Briggs, 1992). An additional comment on the differences in personality measures is noted by Block (1995). He views that lack of agreement by personality researchers regarding the structure of the Big Five is a major flaw in its adoption and measurement. Block argues that further research is needed to specify the composition of personality and states that the unquestioned adoption of the Big Five appears premature. The inconsistent results reported in the present study tend to support Block's contentions.

The third limitation was the measurement of the absenteeism and turnover criteria. For absenteeism, a differentiation was not made between voluntary and involuntary absenteeism. Johns and Nicholson (1982) noted that even within these categories there is degree of variability as to why someone chooses to be absent a particular day. In addition, Hackett and Guion (1985) showed that the test-retest reliability of time lost measures (i.e. a measure of voluntary absenteeism) similar to the one used in this study is .66. This may have explained some of Sample Three's inconsistent results. It is possible that a clearer specification of the type of absenteeism could have increased the correlations between Conscientiousness and illegitimate absenteeism.

Another problem of using a time lost measure of absenteeism as the criterion is that it in some cases it may actually be functional to be absent from work (Johns & Nicholson, 1982; Steers & Rhodes, 1978). For example, the school bus drivers contained in Sample One are not only exposed to a variety of illnesses by children, but also run the risk of exposing their illnesses to children. A similar argument could be made for the transit bus drivers in Sample Two. As such, the personality variables associated with involuntary or legitimate absenteeism may differ from absenteeism used for
withdrawal (Steers & Rhodes, 1978). For example, one could score low on the Emotional Stability dimension and be absent for legitimate reasons and have the same number of absenteeism days as one who is high on the Emotional Stability dimension and low on the Conscientiousness dimension who misses work because of illegitimate reasons. One could make the argument that those individuals who are high on the Conscientiousness dimension are absent because they do not want to expose others to their illness, or are more concerned with caring for a sick child or relative at home. On the other hand, individuals low on the Conscientious dimension may miss work simply because they view absenteeism as an entitlement, or as a means of withdrawal. However, for both high and low Conscientiousness individuals the rate of absenteeism may be the same. Without knowing the reason for absenteeism, and the individual’s motivation for being absent (Martocchio & Judge, 1994), it is difficult to determine with precision the exact nature of the relationship between personality and absenteeism. The low correlations may therefore be a function of using raw absenteeism as a measure, rather than the type of absenteeism (i.e., legitimate versus illegitimate). Future studies should examine the reason for being absent rather than just the occurrences of absenteeism (Hackett & Guion, 1985) to better identify the personality variables associated with each type (Steers & Rhodes, 1978).

Similarly, for the turnover criterion measure, a distinction was not made between voluntary and involuntary turnover for Sample One. Because Campion (1991) suggested that the causes for each are likely to be dissimilar, the low correlations between the personality variables and turnover may be misleading. In addition, although Sample Three distinguished between voluntary and involuntary turnover, this differentiation was made at a gross level. Campion has also suggested that the reasons for voluntary turnover may be impacted differentially by antecedent variables and should therefore be studied at a more micro level of analysis. For example, leaving a position because of a transfer by a spouse may have different causes than leaving a position to accept a better offer of employment even though both are considered voluntary. The role of Conscientiousness may play a more significant role to predict the latter rather than the former.
A fifth limitation of the study, and a potential area for future research, was the assessment of the structural equivalence of the personality and biodata scales. The following evidence would lead one to believe that the biodata scales were not equivalent to the personality scales. Consequently, the use and interpretation of the biodata scales as parallel measures or construct valid measures of the Big Five appeared unjustified. For all three samples it was shown that a large percentage of items that were used for the personality scales were not used for biodata. Moreover, reliabilities for the biodata scored scales differed from the personality scales, and in every case the reliability of the former was lower than the latter. The reliabilities of the biodata scored scales in this study are consistent with previous research and suggest that there are a number of multidimensional constructs are being assessed with a limited number of items (Cortina, 1993; Mumford & Owens, 1987). Evidence from the incremental regression results showing an increase in $R^2$ when biodata scored scales were entered and the different profiles obtained when biodata scoring was used provides additional support for the argument.

Finally, the PAMS profiles were subject to a number of plausible interpretations. Because the profiles did not conform to those specified by the theory, the interpretations were exploratory and subject to alternative interpretations. For Sample Three, profile differences were noted for high versus low absenteeism individuals for each of the six years of data. Unfortunately, these differences were not consistent and did not allow one to make concrete inferences about the personality of a low absenteeism individual. It is uncertain whether these inconsistent results were due to measurement artifacts or a genuine differences without the use of confirmatory methods.

FUTURE STUDIES

Many of the limitations above should be examined in future studies. For example, studies could be conducted that assess the differences between applicant and incumbent samples. Future studies are needed to (a) determine the impact on criterion relationships for the two groups, and (b) provide a better specification of the sixth factor in applicant samples. The existence of a sixth factor in applicant samples makes one question the validity of the Big Five personality model across settings and offer
merit to Block's (1995) assertion that "Serious uncertainties have arisen in regard to the claimed 5-factor structure and the substantive meanings of the factors" (p. 187).

Future investigations should examine the relationship between personality and criterion variables at a finer level of analysis for both the predictor and criterion measures. Because both turnover (Campion, 1991; Jackofsky, 1984) and absenteeism (Farrell & Stam, 1988) occur for a number of reasons beyond a simple voluntary-involuntary dichotomy, an assessment of personality at a facet level of analysis may lead to increased prediction. For example, the HPI facet scale for "No Somatic Complaint" may explain absenteeism due to actual illness; the facet scale for "No Guilt" may explain absence for individuals who take time off work for recreational purposes; the facet scale for "Likes Parties" may explain absences that occur primarily on Mondays or Fridays; and the "Good Attachment" scale may be associated with long-term tenure.

Using PAMS to explore personality profiles at several levels of analysis is also a fruitful area of study. Models could be tested to determine how one's fit to the prototypical profile influences initial withdrawal behaviors (e.g., absenteeism) followed by voluntary turnover over time. These analyses could be done at the occupational, organizational, and departmental level. Additionally, for some high turnover occupations there is an influx of new employees that may alter the initial profile. Studies could be designed to determine whether members of the cohort group leave the organization or alter the composition of the predominant profile.

Although the superiority of biodata scoring received some support in this study, the structure of the empirical scales was not determined. Studies are needed to determine how scoring influences the factor structure of extant personality tests. In addition, this study used the vertical percent method of scoring. It would be interesting to study how alternate test scoring methods (e.g., horizontal percent, phi coefficient, mean response, etc.) impact factor structure.

The influence of biodata scoring within and between organizations and occupations is also worthy of future study. The results of this study suggest that keys tailored to a specific occupational group or
department may produced higher validities than those that are generalized across occupations or
departments. The question is at what cost? Many occupations or departments do not contain the
requisite sample size to develop scales and therefore scales could not be developed. In addition, the
development of specific scales is somewhat labor intensive and developing keys for each department
may be cost prohibitive. As noted by Rothstein et al. (1990) additional work is needed to determine
whether the costs needed to develop specific scales are offset by the potential benefits of improved
validity.

Finally, Mumford and Stokes (1992) stated that many biodata measures are subject to temporal
decay. Additional studies are needed to determine if temporal decay of biodata measures is equivalent
for absenteeism and turnover. In all likelihood, the temporal decay of absenteeism would be greater
than turnover due to the influence of absence cultures and the low reliability of absenteeism measures.
Moreover, scales developed for specific reasons of absenteeism may be less prone to temporal decay
than scales that are not.

PRACTICAL IMPLICATIONS

Results from this study may cause one to question the use of personality measures to predict
absenteeism and turnover. The correlations between personality with turnover and absenteeism were
small and inconsistent for all three samples. Although some of the uncorrected mean correlations
were similar to those reported by Barrick and Mount (1991), most would consider the correlation
coefficient of .13 to be small (Cohen & Cohen, 1983). Nonetheless, the opportunity exists to portray
these results in a positive fashion. But, before presenting the results in a positive light, one must
consider a recent statement by Block (1995) concerning the Big Five (FFA): "...I believed that
advocates of the FFA had, with extraordinary frequency and fervor, presented conclusions in support
of the FFA far stronger than was warranted" (p. 226). As such, results from this study provide
disconfirming evidence that Conscientiousness is associated with absenteeism for two samples, and
small but significant associations in a third. Both samples produced personality-turnover associations
that were not significant. Therefore, organizations that intend to use personality measures to predict absenteeism and turnover should do so with caution.

In addition, results from the personality profile analyses were inconsistent and the profiles were subject to a number of interpretations. For example, a number of plausible explanations could be presented to account for a profile for either someone who remained with an organization, or someone who exhibited low levels of absenteeism. Perhaps the time has come to consider Block’s suggestion to slow the Big Five bandwagon (1995) and investigate alternative models and measures of personality. Further, the existence of a number of different profiles leads one to believe that personality measure may be situationally specific. Because the findings of Cellar et al. (1996) and Schmit and Ryan (1993) indicate that the factor structure of personality measures may differ from sample to sample, one might question the transportability of personality findings across occupations and the application of meta-analytic findings.

Samples One and Two did not support the use of empirical scoring of personality measures to improve prediction. On the other hand, Sample Three supported the use for increasing the association between personality scales and the measures of absenteeism and turnover. The empirical scoring of personality measures appeared to be most useful for applicant samples. Therefore, the findings of Stokes et al. (1993) regarding differences in empirical scoring between applicant and incumbent samples for biodata scales may also apply to the empirical scoring of personality measures. However, this question must be addressed in future research.

Evidence suggests that empirical scoring impacts the meaning and measurement of personality scales. Therefore, future studies are needed to assess the construct equivalence of traditional and empirically scored measures. The profile analyses for biodata scored scales was not useful as the meaning of the scales may have been altered through empirical scoring. These analyses served to underscore the differences between personality and biodata.
CONCLUSIONS

This study demonstrates the need to replicate research. If one were to base conclusions only on Sample Three, the results might be optimistic for using personality measures and empirical scoring. However, the results for the other two samples disconfirmed the adoption of these methods and the use of personality tests for school bus drivers or transit operators. As such, results from these studies posed more questions than they addressed, and one would surmise that additional work is needed before blindly accepting the Big Five as "the" measure of personality.

In addition, this study supports the need to address differences between applicant and incumbent samples for personality and biodata measurement. Therefore, empirical scoring of personality measures may increase prediction but only under very specific conditions. Future studies with alternate samples, different measures of personality, and different, more clearly defined criteria, are needed to determine the utility of using empirical scoring to increase predictive validity of personality measures.

Finally, these results suggest that the situational specificity hypothesis for personality has not been disconfirmed. Future studies are needed to better understand the divergent conclusions from the meta-analytic findings of Tett et al. (1991) and Barrick and Mount (1991) regarding the relative importance of the Big Five personality dimensions.
REFERENCES


## APPENDIX

### Strong’s Tables of New Weights for Differences in Per Cents

<table>
<thead>
<tr>
<th>Part A (To be used when both per cents are between 8 and 92)</th>
<th>Part B (To be used when one per cent is between 3 and 7, or 93 and 97)</th>
<th>Part C (To be used when one per cent is between 0 and 2, or 98 and 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diff. in %</strong></td>
<td><strong>Net Weight</strong></td>
<td><strong>Diff. in %</strong></td>
</tr>
<tr>
<td>69</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td>68</td>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>67</td>
<td>25</td>
<td>67</td>
</tr>
<tr>
<td>66</td>
<td>24</td>
<td>66</td>
</tr>
<tr>
<td>65</td>
<td>23</td>
<td>64 - 65</td>
</tr>
<tr>
<td>64</td>
<td>22</td>
<td>63</td>
</tr>
<tr>
<td>62 - 63</td>
<td>21</td>
<td>62</td>
</tr>
<tr>
<td>61</td>
<td>20</td>
<td>60 - 61</td>
</tr>
<tr>
<td>60</td>
<td>19</td>
<td>58 - 59</td>
</tr>
<tr>
<td>58 - 59</td>
<td>18</td>
<td>57</td>
</tr>
<tr>
<td>56 - 57</td>
<td>17</td>
<td>55 - 56</td>
</tr>
<tr>
<td>54 - 55</td>
<td>16</td>
<td>53 - 54</td>
</tr>
<tr>
<td>52 - 53</td>
<td>15</td>
<td>50 - 52</td>
</tr>
<tr>
<td>50 - 51</td>
<td>14</td>
<td>48 - 49</td>
</tr>
<tr>
<td>48 - 49</td>
<td>13</td>
<td>45 - 47</td>
</tr>
<tr>
<td>45 - 47</td>
<td>12</td>
<td>42 - 44</td>
</tr>
<tr>
<td>42 - 44</td>
<td>11</td>
<td>39 - 41</td>
</tr>
<tr>
<td>39 - 41</td>
<td>10</td>
<td>35 - 38</td>
</tr>
<tr>
<td>36 - 38</td>
<td>9</td>
<td>31 - 34</td>
</tr>
<tr>
<td>33 - 35</td>
<td>8</td>
<td>27 - 30</td>
</tr>
<tr>
<td>29 - 32</td>
<td>7</td>
<td>23 - 26</td>
</tr>
<tr>
<td>24 - 28</td>
<td>6</td>
<td>19 - 22</td>
</tr>
<tr>
<td>21 - 23</td>
<td>5</td>
<td>15 - 18</td>
</tr>
<tr>
<td>16 - 20</td>
<td>4</td>
<td>11 - 14</td>
</tr>
<tr>
<td>12 - 15</td>
<td>3</td>
<td>7 - 10</td>
</tr>
<tr>
<td>8 - 11</td>
<td>2</td>
<td>4 - 6</td>
</tr>
<tr>
<td>3 - 7</td>
<td>1</td>
<td>2 - 3</td>
</tr>
<tr>
<td>0 - 2</td>
<td>0</td>
<td>0 - 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
VITA

The author was born in Fort Knox, Kentucky, on September 24, 1959. He graduated from La Jolla High School in 1977 and attended the University of California at San Diego, and San Diego State University in the early 1970s before entering the restaurant industry. Mr. Davis had a successful career in restaurant management and returned to San Diego State in 1987 to complete his undergraduate degree in psychology. After graduation in 1990, the author attended Louisiana State University where he completed his master's degree in psychology in 1992.

The author left Baton Rouge in 1994, and returned to California in order to collect the data for his dissertation and work for the Human Resources Department at North County Transit District. Mr. Davis is currently employed in the Recruitment and Selection Division of the Department of Human Resources at the County of San Diego. The author's research interests include personality testing, biographical data, person-organization fit, adverse impact analyses, performance appraisal, and turnover.

He is married to Suzanne Marie Davis and the couple are expecting their first child in the summer of 1997.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Bruce W. Davis

Major Field: Psychology

Title of Dissertation: An Integration of Biographical Data and Personality Research Through Sherwood Forest Empiricism: Robbing from Personality to Give to Biodata

Approved:

[Signature]

Major Professor and Chairman

[Signature]

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination:

June 12, 1997