1999

On Biodata Construct Validity, Criterion-Related Validity, and Adverse Impact.

Michelle Ann Dean
Louisiana State University and Agricultural & Mechanical College

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

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/6938

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
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

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ON BIODATA CONSTRUCT VALIDITY, CRITERION-RELATED VALIDITY, AND ADVERSE IMPACT

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 Interdepartmental Program in
Business Administration

by
Michelle Ann Dean
B.S., Louisiana State University, 1992
M.B.A., Louisiana State University, 1994
May 1999
ACKNOWLEDGMENTS

I would like to thank Dr. Dana Broach of the Federal Aviation Administration Aeromedical Research Institute for making available the data analyzed in this dissertation. I would also like to express thanks to my dissertation committee members: Drs. Nate Bennett, Eric Braverman, Paul Jarley, Kevin Mossholder, and Craig Russell for their dissertation input and suggestions as well as their guidance throughout my doctoral studies. I would also like to thank my family for their encouragement, especially my grandmother, Pauleta Dean, who passed away while I was completing my dissertation.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................. ii

ABSTRACT ...................................................................................... iv

CHAPTER
  1  INTRODUCTION ................................................................. 1
  2  REVIEW OF THE LITERATURE ............................................. 15
  3  METHODS .............................................................................. 48
  4  RESULTS ................................................................................. 71
  5  DISCUSSION AND CONCLUSION ........................................ 105

REFERENCES .................................................................................. 128

APPENDIX ..................................................................................... 142

VITA ............................................................................................... 143
ABSTRACT

This study examined the personnel selection technique of biographical information (biodata) in terms of theory, criterion-related validity, and adverse impact. First, the construct validity of biodata was examined to determine if biodata theory was useful in explaining biodata's strong criterion validity. Items from an existing biodata inventory were mapped onto construct domains drawn from Mumford, Stokes, and Owens' (1990) ecology model. Relationships between subjects' biodata responses and training performance was examined for consistency with the model's predictions in an organizational sample. The ecology model did not fit the data well. Follow up exploratory analyses did yield good fit when the model was extended by grouping construct domains within developmental time periods.

Second, biodata was examined in terms of simple and incremental criterion-related validity relative to a general cognitive ability test. The biodata instrument was also investigated in terms of incremental criterion validity of biodata predictor scales used in combination with a general cognitive ability, or "g," test. Predictor scales consisted of all biodata response options, "g-loaded" response options, and "non-g-loaded" response options, respectively. The biodata scale (including all biodata items) outperformed the general cognitive ability test both individually and incrementally (both before and after correcting for the effect of range restriction due to selection on g). The biodata g and non-g item sub-scales slightly outperformed the test of general cognitive ability.
Finally, biodata adverse impact was assessed in two ways. First, individual biodata response options were examined for possible adverse impact. Second, separate biodata scales including and excluding adverse impact response options and a test of general cognitive ability were compared in terms of adverse impact. Eliminating response options that violated the four-fifths rule resulted in a relatively large decline in the standardized mean difference between subgroups, no appreciable decrease in biodata criterion-related validity, and minimal adverse impact relative to both the biodata scale containing all response options and the general cognitive ability measure. Research findings are discussed and implications for theory, future research, and practice are offered.
CHAPTER 1

INTRODUCTION

The argument could be made that people are organizations’ most valuable assets (Beatty, Schneier, & McEvoy, 1987). If this is indeed true, the method by which organizations select employees is critical. Any selection device can be evaluated against the degree to which it:

1) identifies people who will perform best on the job (i.e., maximizes predictive power),
2) complies with Federal regulations on employee selection, and,
3) contributes to development of a theory of performance prediction.

The literature on scored biographical information (hereafter simply biodata), typically gives this selection device high marks with respect to the first and second criterion. Narrative and meta-analytic reviews consistently report average cross-validities between .30 and .40 (Asher, 1972; Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). Reilly and Chao (1982) concluded that only biodata and peer evaluations have criterion validities roughly equal to those reported for general cognitive ability tests. Of equal significance, biodata has been reported to have a low degree of adverse impact (Pace & Schoenfeldt, 1978) compared to levels characteristic of cognitive ability tests (U.S. Employment Service, 1970).
Despite the cumulative evidence of biodata's validity and lack of adverse impact on minority populations, it has not been received with widespread acceptance by practitioners. For example, Hammer and Kleinman (1988) found only 6.8% of 248 firms surveyed had ever used biodata in employment decisions and only 0.4% currently used biodata. Low usage rates among practitioners may be linked to biodata's poor standing within the academic community where it is frequently cited as an example of atheoretical "dustbowl empiricism" (Childs & Klimoski, 1986; Dunnette, 1962; Owens, 1976; Nickels, 1994). Organizations may not be willing to employ a selection device without knowing why it predicts performance.

The purpose of this dissertation was to examine why biodata predicts subsequent work performance and how biodata compared to a test of general cognitive ability in terms of criterion related validity and compliance with Federal regulations regarding adverse impact. The primary goal was to empirically examine predictions derived from current biodata theory. Items from existing biodata inventories were mapped onto construct domains drawn from Mumford, Stokes, and Owens' (1990) ecology model. Relationships between subjects' responses to these items and a training performance criterion were examined for consistency with the model's predictions. A second goal of this research is to estimate the degree to which biodata and general cognitive ability tests individually and incrementally predict performance. The degree of adverse impact was estimated for each individual predictor. Given
that minorities typically score one standard deviation below majority applicants on standardized tests of cognitive ability, the extent to which biodata scales achieve comparable levels of criterion-related validity while minimizing adverse impact becomes an important practical question (U. S. Employment Service, 1970). The remainder of this chapter briefly describes biodata items, biodata theory, and literatures comparing biodata and general cognitive ability criterion-related validities and adverse impact.

**Biodata Overview**

Scored biographical information, or biodata, consists of life history information gathered using paper and pencil self-report questionnaires. Biodata focuses on past life experiences (or their correlates) that are presumed to causally influence personal development which, in turn, influence criterion performance (Owens, 1976). In selection scenarios, candidates responses to questions about prior life experiences are used to predict subsequent criteria (e.g., job performance, turnover, etc.). Items included in a biodata inventory capture developmental life experiences, typically emphasizing either the magnitude or frequency of an experience occurrence. Example items from biodata inventories include:

- On the average, how many hours of homework did you do a week in high school? (Owens & Scheonfeldt, 1979)

- How successful were your teachers in arousing your academic interests? (Owens & Schoenfeldt, 1979)
• How often have you set long term (more than a year) objectives or goals for yourself? (Russell, Mattson, Devlin, & Atwater, 1990)

• How often did you learn about procrastination the hard way? (Russell, Mattson, Devlin, & Atwater, 1990).

Items are usually in multiple choice format and are optimally weighted to predict a criterion of interest (Mumford & Owens, 1984; Owens, 1976).

As noted above, when appropriately scored, biodata inventories are characterized by strong criterion-related validities (Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). While empirical evidence shows strong support for biodata's predictive ability, researchers have expressed concerns for over 35 years that specific theoretical rationales have been slow to surface (Dunnette, 1962; Henry, 1966).

On Biodata Construct Validity

In 1902, Sir Francis Galton stated “the future of each man is mainly a direct consequence of the past—of his own biological history, and of those of his ancestors. It is therefore, of high importance when planning for the future to keep the past under frequent review...” (Galton, 1902; p. 2). This quotation is reflected in the behavioral consistency principle (Wernimont & Campbell, 1968) which holds that the best predictor of future behavior is past behavior. This principle assumes generally consistent behavior within-person, across time (Wernimont & Campbell, 1968) and is an oft used rationale for biodata’s predictive abilities.
Despite the rationale’s intuitive appeal, it is too narrow for situations where no prior behavior exists, thus limiting its usefulness for biodata item development. Biodata is frequently used in scenarios where applicants may not have previous work experience, and therefore, no past behaviors that resemble the desired future behaviors (e.g., Russell, et al., 1990). For biodata item development, the behavioral consistency principle provides a rationale only for those items that tap aspects of the criterion construct domain at previous points in time (Russell, 1994).

Henry (1966) predicted that lack of insight into why biodata predicts may, in the long run, set undue upper bounds on the predictive ability of biodata measures. A more comprehensive rationale was developed by Mumford, Stokes, Owens and colleagues. The ecology model (Mumford & Stokes, 1991; Mumford, Stokes, & Owens, 1990) proposed individuals select themselves into situations based on perceived value of expected situational outcomes. Individuals’ pre-existing intellectual, interpersonal, and social characteristics were expected to influence these choices. Each new situation was hypothesized to require adaptation by the individual and could be viewed as a developmental experience. The ecology model assumed earlier activities and experiences were direct predictors of later individual differences. The model explicitly hypothesized that people develop and change with each new experience. Specific construct domains hypothesized to influence subsequent behavioral outcomes in the ecology model include: social, personality, and
intellectual resources; choice processes (e.g., goals, needs, values, and beliefs); and filtering processes (e.g., locus of control and self-image).

**Theoretical implications for biodata construct validity.** Despite previous efforts to offer a theoretical rationale for biodata selection technology, biodata's dustbowl empiricism label remains. This label may still plague biodata due to continued item development absent clear ties to theory-based construct domains. In a review of the personnel selection literature, Schmidt, Ones, and Hunter (1992) suggested greater attention be paid to constructs underlying biodata items. They argued lack of attention to latent construct domains prevents realization of biodata's full potential. The ecology model provides a good starting point from which to attend to biodata construct validity.

The ecology model provides an initial conceptualization of construct domains tapped by biodata items. Unfortunately, it does not provide strong guidance on how to create items that tap their respective biodata construct domain (Russell, 1994). Biodata research has yet to provide a specific link between theory, item content, and performance measures. No prior research has looked at directly comparing this model's abilities to link item content to job performance criteria (Russell, 1994). This type of research is necessary in order to develop strong theory in biodata.

Greenwald (1975) described strong theory as being both operationally and conceptually disconfirmable. Specifically, he stated:
“When a theory applies to an empirical area in which there are strongly established operational definitions linking theoretical concepts to research procedures, the effect of data disconfirming a prediction is to call into question the theoretical conceptualization underlying the prediction. When operational definitions are not so firmly established, it is a reasonable response of the theorist to interpret unexpected data as calling into question the appropriateness of research operations before abandoning the theoretical conceptualization. When the relation between theory and data is characterized by questionable operations of the latter sort...the theory will be said to be characterized only by operational disconfirmability. When the link between concepts and operations is more confidently established, the theory will be said to be characterized by the stronger level of disconfirmability, conceptual disconfirmability” (p. 494).

Currently biodata research might be characterized as being only operationally disconfirmable due to the lack of 'firmly established' links between latent biodata constructs and biodata items.

**On Biodata Criterion-Related Validity**

Interestingly, the argument could be made that both biodata and general cognitive ability research can be traced back to the work of Sir Francis Galton. Galton (1869) viewed general ability in terms of both biology and evolution. The notion of evolution relates closely to biodata’s underlying rationale—that individuals evolve and change due to situations in which they find themselves or that they consciously choose. Galton provided a conceptual integration of biodata and general cognitive ability research when he suggested general mental ability and life experiences are inextricably interconnected throughout life. This common link was largely ignored by later selection research as the general cognitive ability and biodata research streams took divergent paths.
Specifically, Spearman (1904; 1927) greatly advanced cognitive ability research through the mathematical operations of factor analysis. The approach to mental ability that evolved out of Spearman's work was labeled psychometric g, or simply g (Jensen, 1986). The g construct has been found to be predictive of job performance across almost all jobs (Hunter & Hunter, 1984), especially jobs characterized by high task complexity (Hunter, 1986). In contrast, biodata researchers have focused on prior life experiences which may have played a role in developing intellect. Virtually no investigators pursued both research arenas simultaneously.

As noted above, meta-analyses comparing simple criterion-related validities of different predictors consistently report g and biodata to be valid performance prediction technologies (Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). Mitchell (1996) suggested uneven application of statistical corrections give mental ability tests an edge over biodata criterion validities. He noted typical uncorrected general cognitive ability criterion validity with job performance to be between .15 and .30 and typical biodata cross-validities to range between .30 and .40. Biodata findings are generally not subjected to statistical corrections, unlike general cognitive ability tests, which are typically subjected to numerous corrections. Alternatively, Thorndike (1986) suggested cognitive ability measures may not compare favorably in relation to non-cognitive measures (such as biodata) when selection systems initially screen out those low in g (e.g., an initial hurdle.
requiring a college degree). When the applicant pool is range restricted on g, measures of cognitive ability may not add utility to the selection process.

Biodata and general cognitive ability measures both enjoy high validities and may tap overlapping construct domains. While tests of g do not measure importance or frequency of prior life experiences, biodata items tapping prior life experiences may reflect causal influences that led to the current level of g. Biodata has been labeled a "non-cognitive" predictor due to the absence of "correct" answers to items on biodata inventories. Clearly, this label may be inaccurate. Biodata may capture aspects of both the acquired and innate side of intellectual development. Biodata also capitalizes on situation context and may be useful for determining typical motivation to perform. Regardless, some biodata-g domain overlap may be present. Biodata theory, specifically, the ecology model (Mumford, Stokes, & Owens, 1990) hypothesized that individuals are influenced by their environment and subsequently self-select themselves into situations and experiences maximizing expected outcomes. This interactionist perspective recognized the importance of both person characteristics (e.g., g) and the environment.

In contrast, general cognitive ability tests do not provide information regarding g acquisition or motivation. Instead they are concerned with applying a common metric across persons with no consideration of context (Mitchell, 1996; i.e., measures of g only consider context in terms of attempts to standardize testing settings). It would appear that biodata and tests of g might
augment one another's criterion validity. Interactionists suggest "situations are as much a function of the person as the person's behavior is a function of the situation" (Bowers, 1973, p. 327). Schneider advocated an interactionist position as the most accurate representation of organizational behavior (1983; 1987). Schneider and Schneider (1994) applied Schneider's attraction-selection-attrition (ASA) model (1987) to provide an organizational framework for biodata prediction. The ASA model suggested individuals are attracted to, selected by, and stay with organizations matching their personal characteristics. These cumulative person characteristics in turn defined the organization (i.e., "the people make the place," Schneider, 1987, p. 437). Schneider and Schneider (1994) suggested aggregate life history information be used at the organizational level to determine the types of experiences individuals need in order to achieve individual and organizational effectiveness.

A further distinction between biodata and g criterion validity may be found in the recent distinction between typical and maximal job performance. Firms typically want to select employees with high long-term performance. Yet when individuals take general cognitive ability tests, they know there is a correct answer and their motivational level to perform is probably higher than that exhibited in the typical long-term job performance situation. In contrast, biodata instruments lack an obviously correct answer and hence cannot encourage the test taker to exhibit transitory maximal performance.
Interestingly, Sackett, Zedeck, and Fogli (1988) reported maximal and typical job performance were only modestly correlated (.16 and .36 in new employee and current employee samples, respectively). This finding suggests typical and maximal job performance measures tap different latent construct domains and, in turn, may have different antecedent causal influences and predictors. Again, biodata may be more useful in predicting typical motivation levels.

In sum, evidence and theory suggest biodata inventories and general cognitive ability tests tap overlapping but non-identical predictor construct domains. The exact implications of this observation for criterion validity inferences remains to be seen.

**On Biodata Adverse Impact**

Selection methods in compliance with the Equal Employment Opportunity Commission's (EEOC) *Uniform Guidelines on Employee Selection Procedures* (1978) are generally viewed favorably in any subsequent litigation. Organizations are faced with the dilemma that general cognitive ability tests predict job performance while exhibiting adverse impact on racial and ethnic minority groups. Adverse impact occurs if a test causes employers to reject a larger proportion of minority than majority applicants. The Equal Employment Opportunity Commission has made clear that when conducting validation studies, employers should consider available alternatives which will achieve their legitimate business purpose with lesser adverse impact. Employers
cannot rely solely on establishing the validity of an instrument or procedure. (EEOC Uniform Guidelines, 1978). General cognitive ability tests consistently report mean race differences of up to one standard deviation between the majority and minority groups (U.S. Employment Service, 1970), which necessarily result in adverse impact when such tests are used for selection.

Differences in standardized general cognitive ability scores by race are reported starting in early childhood. Studies report Caucasian and African-American students' standardized test scores were nearly identical at the beginning of schooling, but performance gaps started to appear soon after (Alexander & Entwhistle, 1988; Burton & Jones, 1982). Gerard (1983) reported that sixth grade Caucasian children outscore Black counterparts by two full grade levels. Various explanations have been given for this phenomenon including differences in socioeconomic status, culture, and genetic make-up (Herrnstein & Murray, 1994).

A major advantage of biodata is its ability to select applicants with minimum adverse impact. This is achieved through empirically-derived biodata scoring keys. Very simply, adverse impact is eliminated by deleting response options from the key that demonstrate differential criterion prediction across protected groups (e.g., race, gender). Pace and Schoenfeldt (1977) recommended that for purposes of maximizing compliance under the EEOC Uniform Guidelines (1978), biodata item development should be guided by job analysis and overall test score and each item be individually examined for
adverse impact. The EEOC Uniform Guidelines (1978) only require that the overall test score be examined for adverse impact.

As the workforce continues to become more diverse, organizations are more likely to face tradeoffs between maximizing validity versus maximizing organizational diversity. The continued concern that cognitive ability tests unfairly screen out minorities from selection consideration (Gottfredson, 1986) coupled with increasing workforce diversity has made alternative prediction techniques with less adverse impact, such as biodata, more attractive.

Research Purpose

This chapter introduced three key goals of selection research: 1) developing theoretical rationales for performance prediction, 2) maximizing predictive power, and 3) minimizing adverse impact. This investigation examined biodata in the context of these goals. Specifically, this research investigated why biodata predicts subsequent work performance and compared biodata and general cognitive ability tests’ individual and incremental criterion-related validity. Adverse impact of biodata instruments and a general cognitive ability measure were compared. Examination of why biodata demonstrates criterion validity involved empirically examining predictions derived from current biodata theory. Items from an existing biodata inventory were mapped onto construct domains drawn from Mumford, Stokes, and Owens’ (1990) ecology model. Relationships between subjects’ responses to these items and training performance were examined for consistency with the model’s predictions in a
sample of air traffic controllers hired by the U.S. Federal Aviation Administration (FAA) from 1986 to 1992. In addition, analyses were performed to estimate the degree to which biodata and general cognitive ability tests predicted performance both individually and incrementally. Importantly, the degree of adverse impact will be estimated for biodata and general cognitive ability measures.
CHAPTER 2

REVIEW OF THE LITERATURE

As early as 1894 biodata was used as a personnel selection device when life insurance agents were asked a standard set of questions about previous life experiences. Question content included topics such as applicants' past insurance sales experience and number of places lived in the past decade (Ferguson, 1961). Biographical data evolved out of the use of job application blanks and variants such as the weighted application blank (WAB, England, 1961). Typical weighted application blanks primarily included demographic items (e.g., age, years of education, previous occupations, and marital status) and weighted this information according to each item's ability to differentiate between successful and non-successful employees. Modern biodata items differ, tapping a wider range of prior behavior and experiences than simple demographic information (Mumford & Owens, 1987).

The majority of biodata research over the past century has been motivated by one of two objectives: 1) maximizing predictive validity and 2) understanding why biodata predicts. The bulk of work on biodata has focused on the former through examination of varying biodata scoring procedures (e.g., Devlin, Abrahams, & Edwards, 1992; Mitchell & Klimoski, 1982). Emphasis on maximizing criterion-related validity may have contributed to biodata's "dustbowl empiricism" stigma (Childs & Klimoski, 1986; Dunnette, 1962; Nickels, 1994; Tenopyr, 1994). Other possible explanations for biodata's

15

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
A theoretical reputation may be researchers failure to link item content and development to construct domains explicated in biodata theory (Russell, 1994). Additionally, mild controversy over what should be classified as biodata may have hampered theory development (Asher, 1972; Henry, 1966; Mael, 1991).

The primary goals of the current research effort are to 1) investigate why biodata predicts subsequent work performance by empirically examining predictions derived from current biodata theory, 2) compare biodata and general cognitive ability tests' individual and incremental criterion-related validity, and 3) compare the degree of adverse impact exhibited by biodata instruments and general cognitive ability measures. Literatures directly bearing on these goals are reviewed below including: 1) current biodata theory and construct domains, 2) criterion-related validity and incremental validity of biodata when used in concert with general cognitive ability (g) measures, and 3) adverse impact of biodata and g measures. Research questions examined in this study are listed as developed.

**Biodata Conceptual Domain**

It is necessary to first attempt to define what biodata is prior to empirically examining the selection technique. Previous taxonomic research attempting to define biodata's conceptual domain and general attributes is reviewed and a general definition of biodata is presented. Next, the behavioral consistency principle (Wernimont & Campbell, 1968), the most frequently used rationale for biodata, and the ecology model (Mumford, Stokes, & Owens,
1990), the most recently developed biodata theory, are explicated. The ecology model framework is presented as a theoretical basis for this research.

There is no universally agreed upon definition on what constitutes biodata. Owens (1976) originally suggested biodata items should have a demonstrated or presumed relationship with "personality structure, personal adjustment, or success in social, educational, or occupational pursuits" (p. 613). Some researchers have narrowed the focus of what constitutes biodata to include only historical experiences that are verifiable, while others have more broadly defined biodata to include any items that tap personality, motivation, aspiration, attitudes, and values (Asher, 1972). Many researchers have used the label "biodata" loosely. Some studies have labeled simple demographic information as biodata. With such a wide range of research labeled biodata, this could call into question the generalizability of research findings.

Mael (1991) summarized previous taxonomic work in biodata (Asher, 1972, Cascio, 1982) and grouped biodata attributes into three general categories: 1) historical, 2) methodological, and 3) legal/moral. These taxonomic efforts pinpoint the types of items typically found on a biodata instrument and attempt to differentiate biodata items from other closely related, but conceptually distinct, measures such as personality scales. There is no theoretical basis for "correctness" of any of these attributes. Taxonomies can
useful for generating theory, but should not be considered theory itself (Bacharach, 1989).

Mael (1991) suggested the historical nature of items constitute biodata's defining characteristic primarily due to biodata's reliance on the behavioral consistency principle (Wernimont & Campbell, 1968). Biodata items concern actual past events that have taken place in one's life and do not include hypothetical scenarios such as found in situational interviews (Latham, Saari, Pursell, & Campion, 1980). Situational interviews present candidates with a possible scenario and ask their likely behavioral response. Mael (1991) suggested questions of general attitudes (e.g., Would you describe yourself as shy?) that do not relate to a specific past event are outside the realm of biodata and are more closely aligned with dispositional measures.

Researchers suggested certain methodological attributes aid in obtaining accurate biodata responses (Mael, 1991). Some researchers recommend biodata items be externally focused, objective, first-hand, and verifiable (Asher, 1972; Mael, 1991). Externally focused items tap some action or event in which an individual was involved and are not merely opinions or reactions to an event. Many researchers advocate biodata items be objective rather than subjective. Mael suggested biodata items should ask for respondent's first-hand knowledge, avoiding asking individuals about how others would evaluate the respondent. For example, asking, "How did your parents evaluate your academic achievement?" would be second-hand
information on which the respondent is asked to speculate. An additional recommended methodological attribute is that items be verifiable.

These methodological attributes (externally focused, objective, first-hand, and verifiable) are believed to result in honest, accurate responses by discouraging socially desirable responses or faking. Verifiable items have been found to reduce the effects of faking (Atwater, 1980; Cascio, 1975; Mosel & Cozan, 1952). Hough, Eaton, Dunnette, Kamp, and McCloy (1990) found simply warning respondents that their answers can be verified may act as a faking deterrent.

Mael's (1991) taxonomy also considered legal and moral issues surrounding biodata use. Items may vary in terms of controllability, accessibility, and visible job relevance. Controllability refers to the degree to which a person chose to perform or not to perform an action. This label mimics Owens and Schoenfeldt's (1979) "prior behaviors" (behaviors in which a person chooses to engage; e.g., playing sports in high school) and "input variables" (events or circumstances that happen to a person beyond that person's control, e.g., parents socioeconomic status). Mael suggested all life events, whether consciously chosen or not, have the ability to shape a person's future behavior and should be included on a biodata instrument.

There is some opposition to including non-controllable items such as parental behavior and socioeconomic status, due to applicants' lack of control over their early environment. This leads to a further legal question regarding
whether items should tap skills and experiences that are equally accessible to all applicants (Stricker, 1987; 1988). Stricker (1988) suggested items asking about experience as a football team captain would be unfair because individuals of a particular gender, size, or size of school may not have had the opportunity to engage in this role. Mael (1991) suggested the concept of equal access is irrelevant, rather what is relevant is that the person who had access to the role was changed in some way by the role while others who were not in-role received no benefit nor harm (i.e., individuals who were not football captains are not penalized for non-exposure).

Finally, biodata items vary in visible job relevance. Most researchers view all life experiences as potentially developmental. However, an exception involved Gandy, Outerbridge, Sharf, and Dye's (1989) use of only items with a point-to-point relationship with the job on their public sector biodata instrument in order to avoid the accessibility issue.

Research has examined the influence of item attributes on biodata results. Shaffer, Saunders, and Owens (1986) found "soft," or subjective, non-verifiable items nearly as predictive and reliable as "hard," or verifiable, factual items. Average test-retest reliability five years after initial administration was higher for objective than subjective items. Barge (1987) analyzed biodata items in terms of three dimensions: 1) item heterogeneity (degree to which items measure more than one construct), 2) behavioral discreteness (degree to which items address "a single, perhaps verifiable behavior rather than a more
abstract or summary characteristic" (pp. 3-4), and 3) behavioral consistency
(degree of congruency between the content domain of the biodata item and the
content domain of the target job, i.e., the degree to which an item is a sign
versus sample of behavior; Wernimont & Campbell, 1968). Barge analyzed
103 items taken from Owens' Biographical Questionnaire (BQ) along these
dimensions and found more homogenous items (in terms of consistency and
discreteness) were more valid. Barge's (1987) study is noteworthy because it
was among the first studies to evaluate the impact of item characteristics on
validity (Stokes & Reddy, 1992).

In sum, taxonomic work on biodata attributes is not based on theory but
is merely a summary of previous work that has been labeled "biodata." No
consensus on what represents biodata has been reached. Hence, the
definition for biodata that will be used in this study is: any item aimed at life
experiences which have occurred in the past (Russell, 1997). It is important
that items be historical because this is the only attribute that can truly be
inferred from the term "biographical." Consistent with Mael's taxonomy and the
majority view regarding biodata's defining characteristics, this study will limit its
focus to items that deal with past life events. There is currently no theory that
speaks to the necessity of a particular attribute, such as why a biodata item
must be verifiable or be first-hand. Agreement may be forthcoming if theory-
based rationales for particular attributes are developed. At that point, standard
psychometric procedures (e.g., factor analysis) can be applied to assess
construct validity of these "required" attributes. Until these categorization systems are supported by strong theoretical rationales, they (like other taxonomic categories) must be considered arbitrary (Russell, 1997).

**On Biodata Construct Validity**


The behavioral consistency principle reflects the most commonly used rationale for biodata—that the best predictor of future performance is past performance (Wernimont & Campbell, 1968). Owens (1976) reiterated this by stating that "one of our most basic measurement axioms holds that the best predictor of what a man (sic) will do in the future is what he has done in the past" (p. 625). This principle assumes generally consistent behavior within-person, across time. Despite the rationale's intuitive appeal, it is inapplicable in situations where no prior behavior exists, thus limiting its usefulness for
guiding biodata item development (Dean, Russell, & Muchinsky, in press; Russell et al., 1990).

Specifically, biodata generally does not involve literally predicting future performance from measures of identical past performance. This can be explained by Wernimont and Campbell's (1968) distinction between "samples" versus "signs" of behavior. Samples represent past behaviors used to predict future behaviors drawn from a single common performance domain. Signs are not equivalent to criterion domain behaviors, but are instead drawn from domains hypothesized to 1) causally influence subsequent performance or 2) be highly correlated with those causal influences. The behavioral consistency principle provides a rationale only for those biodata items that tap aspects of the criterion construct domain at previous points in time (i.e., behavioral samples), while in practice, biodata instruments generally use both signs and samples of past behavior to predict future performance outcomes (Russell, 1996).

Biodata is often used in scenarios where applicants may not have previous work experience, and therefore, no past behaviors which resemble desired future behaviors (Russell et al., 1990). For example, if the behavior of interest is ability to sell life insurance, the behavioral consistency principle would only provide a rationale for those items that asked applicants about their previous life insurance sales experience (i.e., a sample of past behavior). However, predictors should not be limited to samples because signs may help
provide insight into the desired behavior. Given "sign" items are drawn from a domain that is not identical to the performance domain, items tapping previous customer service experience may, for example, contribute to biodata predictive power.

The behavioral consistency principle as applied to biodata items does not address why signs predict which is unfortunate because signs can be particularly useful for applicants with no generalizable work experience. Russell et al. (1990) faced this scenario when developing an instrument to predict performance of high school applicants into the U. S. Naval Academy. In that study applicants had no opportunities to exhibit "samples" of Naval Officer behaviors up to that point in their lives, making it necessary to find signs from adolescent and pre-adolescent experiences that might predict future success. Items focusing on school, social, and employment experiences were used to predict Naval Academy success. Russell et al. found these experiences resulted in accurate prediction of subsequent academic and non-academic performance criteria. Further, the empirically keyed biodata scales demonstrated incremental criterion-related validity when combined with measures of general cognitive ability.

Mumford, Owens, and Stokes (Mumford & Owens, 1987; Mumford & Stokes, 1991: Mumford, Stokes & Owens, 1990) greatly refined and extended the consistency principle through development of the ecology model. The ecology model acknowledged that individuals have their own hereditary and
environmental "baggage" that determine initial individual differences, focusing specifically on how individual difference characteristics shape the choices individuals make. The ecology model was developed from Owen's developmental-integrative model (Owens, 1968, 1971; Owens & Schoenfeldt, 1979), which initially proposed that biodata items needed to capture prior behaviors and experiences that affect personal development on individual difference characteristics (e.g., knowledge, skills, and abilities). These individual differences were hypothesized to subsequently affect a person's performance on organizational criteria of interest.

The ecology model considered both individual differences and processes that motivate and influence choices individuals make as predictors of future performance. Specifically, the model suggested people select themselves into situations based on the value of expected outcomes and pre-existing individual difference characteristics. Each choice requires adaptation to new situations and represents a developmental experience. The model represents an iterative process of choice, development, and adaptation. People are constantly faced with making choices and over time will tend to develop characteristic patterns of choices and behaviors.

Mael (1991) suggested the ecology model was most useful as a rationale for items dealing with behaviors and experiences individuals actually choose to engage in, which subsequently develop knowledge, skills, and abilities. Job performance is commonly held to depend on individuals'
knowledge, skills, abilities, motivation, and other personal characteristics (KSAOs: Dunnette, 1966; Fleishman & Quaintance, 1984). The ecology model argued that "life events indicating successful engagement in activities requiring the application of KSAOs similar to those required on-the-job might prove to be useful predictors" (Mumford & Stokes, 1992, p. 81) as well as those events that play a role in developing KSAOs.

Nickels (1990) identified a framework of characteristics and individual differences posited to influence performance later in life in an early attempt to understand dimensions underlying the ecology model. Nickels suggested the lack of understanding of biodata's ability to predict stemmed from the lack of an empirically testable nomological network upon which to develop biodata measures (Nickels, 1990). Nickels reviewed over one hundred and fifty citations of individual differences and known predictive relationships between past behavior or experience and later performance, yielding a preliminary list of 500 possible dimensions. The 500 dimensions were subjected to a series of reviews by subject matter experts to obtain a more manageable and interpretable number of dimensions. A dimension was excluded from further investigation based upon a consensus decision that the dimension...

"a) demonstrated an obvious content overlap with another dimension (e.g., gregariousness and sociability); b) could not feasibly be rated given the information provided by background data items (e.g., attractiveness); c) was inappropriate with respect to the population (e.g., paranoia in a normal population); or d) seemed unlikely to influence the life history of individuals in adolescence and young adulthood" (Nickels, 1990, pp. 28-29).
This process resulted in dimensions being reduced in number from 500 to 44. Five general dimensional categories emerged, three capturing general categories of individual differences posited to influence subsequent performance: personality resources, interpersonal (social) resources, and intellectual resources. Two other categories covered motivation and beliefs or attitudes, labeled choice and filter processes, respectively.

Nickels' study was one of the first attempts to operationalize the ecology model. Mumford, Stokes, and Owens, the primary ecology model architects, subsequently elaborated this framework (see Figure 2.1), changing some of the labels though not the substance of Nickels dimensions (Mumford & Stokes, 1992). The ecology model suggested individual difference constructs "facilitate the attainment of desired outcomes while conditioning future situational choice by increasing the likelihood of reward in certain kinds of situations" (Mumford & Stokes, 1991, p. 81). The first three categories (personality, social, and intellectual resources) are personal characteristics posited to influence future behavior and decisions. The remaining two categories are motivational variables that might affect situational selection and resource application (i.e., choice and filter processes; Mumford & Stokes, 1991; Nickels, 1990).

**Personality Resources.** Nickels suggested this category represented "stylistic or emotional attributes thought to impact effective environmental interactions" (1990, p. 29) such as adaptability, emotional stability, and
Figure 2.1
Ecology Construct Categories

persistence. These resources closely resemble the "Big Five" personality constructs (e.g., Barrick & Mount, 1991). The five factors are commonly labeled: extraversion (e.g., sociable, assertive, ambitious), emotional stability (e.g., secure, anxious, well-adjusted), conscientiousness (e.g., dependable, efficient, achievement oriented), and openness to experience (e.g., cultured, curious, broad-minded; Barrick & Mount, 1991; Mount, 1997). These five
constructs emerge consistently across longitudinal studies, raters, personality inventories, and protected subgroups (Digman, 1990; Mount, 1997).

Sample biodata items from the Federal Aviation Administration (FAA) biodata instrument that may tap personality resources include:

Item 1: During my years in high school, I was singled out for disciplinary reasons:
   a. 5 or more times
   b. 3 or 4 times
   c. twice
   d. once
   e. never

Item 2: In the three years immediately before accepting my first job in my present job series, the number of different full- or part-time jobs I applied for was:
   a. none
   b. 1-2
   c. 3-4
   d. 5-6
   e. 7 or more

Item 1 may tap the personality dimension of emotional stability and degree of adjustment. The fewer times an individual was singled out for disciplinary reasons, the more well adjusted that individual may be. Item 2 may reflect the personality construct of extroversion, ambition, or persistence. The greater the number of jobs applied for may suggest degree of individual persistence and ambition.

Social Resources. Nickels posited social resources might influence effectiveness of interpersonal relations and therefore play a role in situation selection and subsequent behavior/performance. Some example constructs
include self-monitoring, dominance, empathy. Mael's social identity theory speaks to the influence of group membership on one's own personal identity. It could be argued that the more group memberships held, the greater one's interpersonal adeptness. Greater number of group memberships may suggest individual effectiveness in self-monitoring and ability to adjust behavior to match that of the group.

Example biodata items from the FAA biodata instrument that may tap the social resources dimension include:

Item 3: Relative to the other high school students in my major field of study, my classmates would most likely describe my interpersonal skills as:
   a. superior
   b. above average
   c. average
   d. below average
   e. don't know

Item 4: The number of college clubs and organized activities (band, newspaper, etc.) in which I participated was:
   a. 3 or more
   b. 2
   c. 1
   d. didn't participate
   e. didn't go to college

Item 3 asks the respondent to describe his/her interpersonal skills relative to his/her peers in the past. Item 4 may reflect one's interpersonal skill by determining how many social organizations the individual has been involved with in the past.
**Intellectual Resources.** Intellectual resources represent attributes that enable knowledge assimilation and retention affecting one’s ability to make choices and perform efficiently and effectively. An example of an underlying construct biodata items might capture is general cognitive ability, or "g." The ecology model refers to g but does not speak to the iterative, bi-directional relationship of g to life events and performance over time. Regardless, many biodata items seem to tap g. For example, questions on Owens’ (1971) Biographical Questionnaire asked individuals about academic achievement, academic attitude, and intellectualism. Biodata instruments may tap g in items capturing past experiences requiring or aiding in the development of general cognitive ability. In contrast, paper and pencil tests of general cognitive ability infer g from frequency with which individuals select factually correct answers to questions tapping various knowledge content domains, where there is some universal agreement as to the correct answer (e.g., "4" is the correct answer to the arithmetic knowledge question, "What is 2 plus 2?").

Example biodata items from the FAA biodata instrument that may tap the intellectual resources dimension include:

**Item 5:** My class standing in high school put me in the:
   a. top 10%
   b. top 33%
   c. top 50%
   d. top 90%
   e. did not graduate from high school

**Item 6:** My previous supervisors (or teachers if not previously employed) would most likely describe my ability to recall
facts and details of information as:
  a. superior
  b. above average
  c. average
  d. below average
  e. don't know

Item 5 might be a surrogate measure of general cognitive ability via high school grades earned. The higher one's academic standing, the more intelligent that person is presumed to be. Item 6 reflects intelligence through the ability to accurately retrieve information from memory.

Choice Processes. The choice processes domain represents "differential motivational influences with respect to individual differences in performance" (Nickels, 1990, p. 43). Example traits dealing with motivational issues include goal orientation, personal performance standards, and desirability of the reward (e.g., "valence" in expectancy theory terminology; Vroom, 1964). Research on performance prediction suggests high performing individuals must have motivation and ability to perform (Campbell, Dunnette, Lawler, & Weick, 1970). Gottfredson (1997) suggested biodata items capture motivational components of task performance better than paper and pencil mental ability tests. While general cognitive ability measures may best estimate what applicants "can do" measures not specifically targeting general cognitive ability (i.e., "non-cognitive" measures) such as biodata may best estimate what applicants "will do." Some investigators suggested high
criterion-related validities may be partially due to biodata’s ability to tap both ability and motivational construct domains (Mael, 1991; Mitchell, 1996).

Example biodata items from the FAA instrument that may tap the choice processes dimension include:

Item 7: The number of times I elected non-required college math courses was:
   a. 3 or more
   b. 2
   c. 1
   d. Never
   e. Didn’t go to college

Item 8: The number of high school clubs and organized activities (such as band, newspaper, etc.) In which I participated was:
   a. 4 or more
   b. 3
   c. 2
   d. 1
   e. Didn’t participate

Both these items query respondents on participation in voluntary events. Engaging in these events (non-required coursework, extra-curricular activities) may be indicators of motivation.

Filter Processes. Nickels suggested this category represents values, beliefs, and attitudes which may influence self-perception and consequently decisions an individual makes. Constructs in this category included self-esteem, self-efficacy, and locus of control.

Example biodata items from the FAA biodata instrument that may tap filter processes include:
Item 9: My high school teachers would most likely describe my self-discipline as
  a. superior
  b. above average
  c. average
  d. below average
  e. don't know

Item 10: My peers would probably describe me as being:
  a. much more confident than most
  b. somewhat more confident than most
  c. about as confident as anyone else
  d. somewhat less confident than most
  e. much less confident than most

Item 9 may tap an individual's attitude or belief toward self-discipline. If other's viewed an individual as a highly disciplined person in the past, under the behavioral consistency principle, that person will most likely continue to display discipline in the future. Item 10 may reflect self-esteem or global self efficacy. Displaying an air of confidence is a typical outward sign of an individual's self perception. How an individual feels he or she is perceived by others may play a role in shaping the individual's own perception of oneself.

The items used in this analyses are archival and were not developed with ecology model dimensions in mind. Hence, some items may map multiple constructs. Other items may tap constructs other than those identified in the ecology model. Regardless, a possible strength of biodata items is an ability to capture multiple constructs (Mumford & Owens, 1987; Russell, 1994).

A primary goal of this research was to determine whether these items could be mapped onto the framework presented in Figure 1-1, and if so, to
determine whether the framework was useful in explaining biodata predictive ability. Russell (1994) suggested there is a gap between theories of life history and operationalizations (i.e., biodata item content), thus warranting the current analyses. Additional issues must be addressed to evaluate and strengthen existing biodata theory, thus serving to guide future item development. Importantly, research must address whether items grounded in biodata theoretical models result in higher predictive validities than items not explained by theory. Research suggests items developed on the basis of specific theory-based hypotheses are more likely to produce significant relationships with external criterion performance measures compared to those items generated absent theory-based rationales (Kavanagh & York, 1972; Mumford & Owens, 1987; Nickels, 1990; Quaintance, 1981; Williams, 1961). Nickels (1990) posited theory-based items "are more likely to capture differences in the relevant patterns of antecedent events responsible for the predictive power of the biodata item" (p. 22). In light of the Nickels (1990) and Mumford and Stokes (1992) ecology model framework, the following research questions were developed:

**Research Question #1:** Do items display psychometric characteristics (e.g., content validity, item factor analytic loadings, internal consistency reliability) consistent with theory-based construct domains?
Research Question #2: Do relationships among scale scores derived for latent biodata constructs yield convergent and discriminant validities consistent with theory-based construct domains?

Research Question #3: Do items sorted into theory-based construct domains demonstrate higher criterion-related validity than non-theory-based items?

On Biodata Criterion-Related Validity

As noted earlier, the majority of work in biodata focuses on maximizing predictive efficiency. An interesting issue receiving little attention is the individual and combined effects of g and non-g measures on criterion validity and adverse impact. Ability measures present individuals with an immediate, and artificial problem-solving situation intended to demonstrate maximum performance (Mumford & Stokes, 1992). Individuals' problem solving performance is then used to infer performance potential on other tasks. Further, biodata instruments ask individuals to recall their typical behavior in past life events and experiences. Mumford and Stokes (1992) suggested that as a result, biodata measures cannot provide an upper bound prediction of performance potential, but may be useful in predicting observed, or typical performance. Biodata may be more closely related to measures of practical intelligence (Wagner & Sternberg, 1985) than general cognitive ability measures. Biodata and general cognitive ability measures should be more
highly related to the degree to which general cognitive ability was influenced by prior developmental life events and experiences tapped by the biodata measure (Mumford & Stokes, 1992).

Several reviews document biodata inventory criterion validity (Asher, 1972; Hunter & Hunter, 1984; Owens, 1976; Reilly & Chao, 1982; Schmitt et al., 1984). Previously reported meta-analytic results of biodata and g criterion validity are presented in Tables 2.1 and 2.2.

Reilly and Chao (1982) reported average biodata criterion validities for tenure, training, ratings, productivity, and salary ranging between .32 - .46, with an average validity across all criteria of .35. Hunter and Hunter (1984) meta-analyzed many prior findings and provided numerous mean validities for biodata and g. In a re-analysis of Reilly and Chao's (1982) meta-analysis, Hunter and Hunter (1984) obtained mean biodata and g criterion validities of .34 and .38, respectively. Hunter and Hunter (1984) reported the average validity of cognitive ability in the Dunnette (1972) study to be .45. Meta-analyzing military studies, Hunter and Hunter (1984) found average validities for biodata and g to be .20 and .21, respectively. For entry level jobs, Hunter and Hunter (1984) reported mean criterion validities for biodata and g of .37 and .53, respectively. Hunter and Hunter subjected g validities to corrections for numerous artifacts (e.g., sampling error, error of measurement, range restriction, or criterion unreliability) as per Schmidt and Hunter (1977) and
Table 2.1
Meta-analytic Biodata Criterion Validities

<table>
<thead>
<tr>
<th>Study</th>
<th>Criterion</th>
<th>k</th>
<th>$\Sigma N_i$</th>
<th>$r^c$</th>
<th>$\sigma_r^2$</th>
<th>$\sigma_e^2$</th>
<th>$\sigma_p^2$</th>
<th>$\rho^g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reilly &amp; Chao (1982)</td>
<td>Tenure</td>
<td>13</td>
<td>5721</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>3</td>
<td>569</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>Ratings</td>
<td>15</td>
<td>4000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>Productivity</td>
<td>6</td>
<td>661</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>Salary</td>
<td>7</td>
<td>680</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.34</td>
</tr>
<tr>
<td>Schmitt, Gooding, Noe, &amp; Kirsch (1984)</td>
<td>Performance ratings</td>
<td>29</td>
<td>3998</td>
<td>.317</td>
<td>.0357</td>
<td>.0059</td>
<td>.0298</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Turnover</td>
<td>28</td>
<td>28,862</td>
<td>.209</td>
<td>.0144</td>
<td>.0009</td>
<td>.0136</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Achievement/grades</td>
<td>9</td>
<td>1744</td>
<td>.226</td>
<td>.0784</td>
<td>.0047</td>
<td>.0738</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Productivity</td>
<td>19</td>
<td>13655</td>
<td>.203</td>
<td>.0036</td>
<td>.0013</td>
<td>.0023</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Status change</td>
<td>6</td>
<td>8008</td>
<td>.332</td>
<td>.0014</td>
<td>.0006</td>
<td>.0009</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Wages</td>
<td>7</td>
<td>1544</td>
<td>.525</td>
<td>.0157</td>
<td>.0024</td>
<td>.0133</td>
<td>-</td>
</tr>
<tr>
<td>Hunter &amp; Hunter (1984)</td>
<td>Supervisor ratings</td>
<td>12</td>
<td>4429</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>Promotion</td>
<td>17</td>
<td>9024</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>Training success</td>
<td>11</td>
<td>6139</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>Tenure</td>
<td>2</td>
<td>2018</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.00</td>
</tr>
<tr>
<td>Hunter &amp; Hunter (1984) re-analysis of Dunnette (1972)</td>
<td></td>
<td></td>
<td>115</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.38</td>
</tr>
</tbody>
</table>
(Table 2.1 continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Criterion</th>
<th>k</th>
<th>(\sum N_i)</th>
<th>(\bar{r})</th>
<th>(\sigma_r^2)</th>
<th>(\sigma_e^2)</th>
<th>(\sigma_p^2)</th>
<th>(\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suitability</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>All ratings</td>
<td></td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
</tbody>
</table>

- **a** Number of studies in the meta-analysis
- **b** Sum of all sample sizes across all studies
- **c** Average correlation across all studies weighted by sample size
- **d** Observed variance of the correlation coefficients across all studies
- **e** Expected variance due to sampling error
- **f** Variance attributed to true rho across studies
- **g** Population correlation
- **h** Global ratings, suitability, and all ratings were corrected for error of measurement using a reliability of .60
Table 2.2
Meta-analytic General Cognitive Ability Criterion Validities

<table>
<thead>
<tr>
<th>Study</th>
<th>Criterion</th>
<th>k</th>
<th>(\Sigma N_i)</th>
<th>(\bar{r})</th>
<th>(\sigma^2)</th>
<th>(\sigma^2_e)</th>
<th>(\sigma^2_I)</th>
<th>(\rho^g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schmitt, Gooding, Noe, &amp; Kirsch (1984)</td>
<td>Across all criteria</td>
<td>53</td>
<td>40230</td>
<td>.248</td>
<td>.0191</td>
<td>.0012</td>
<td>.0179</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Performance ratings</td>
<td>25</td>
<td>3597</td>
<td>.220</td>
<td>.0156</td>
<td>.0063</td>
<td>.0093</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Turnover</td>
<td>8</td>
<td>12449</td>
<td>.141</td>
<td>.0188</td>
<td>.0006</td>
<td>.0182</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Achievement/grades</td>
<td>5</td>
<td>888</td>
<td>.437</td>
<td>.0221</td>
<td>.0037</td>
<td>.0184</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Status change</td>
<td>9</td>
<td>21190</td>
<td>.282</td>
<td>.0088</td>
<td>.0004</td>
<td>.0084</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Work Sample</td>
<td>3</td>
<td>1793</td>
<td>.426</td>
<td>.0066</td>
<td>.0011</td>
<td>.0055</td>
<td>-</td>
</tr>
<tr>
<td>Hunter &amp; Hunter (1984)</td>
<td>Entry-level job selection (with training after hiring)</td>
<td>425</td>
<td>32,124</td>
<td>.53</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Promotion or certification (where current job performance is basis of selection)</td>
<td>425</td>
<td>32,124</td>
<td>.53</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.15</td>
</tr>
<tr>
<td>Re-analysis of Dunnette (1972)</td>
<td>-</td>
<td>-</td>
<td>215</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.45</td>
</tr>
<tr>
<td>Re-analysis of Vineberg &amp; Joyner (1982)</td>
<td>Supervisor ratings</td>
<td>-</td>
<td>101</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.21\textsuperscript{h}</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>-</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.35\textsuperscript{h}</td>
</tr>
<tr>
<td></td>
<td>Supervisor ratings</td>
<td>-</td>
<td>112</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.28\textsuperscript{h}</td>
</tr>
</tbody>
</table>

\(a\) Number of studies in the meta-analysis
\(b\) Sum of all sample sizes across all studies
\(c\) Average correlation across all studies weighted by sample size
\(d\) Observed variance of the correlation coefficients across all studies
\(e\) Expected variance due to sampling error
\(f\) Variance attributed to true rho across studies
\(g\) Population correlation
\(h\) Global ratings, suitability, and all ratings were corrected for error of measurement using a reliability of .60
Hunter, Schmidt, and Jackson (1982), while biodata validities were only corrected for sampling error.

Schmitt et al. (1984) meta-analyzed all predictor criterion validities reported in *Journal of Applied Psychology* and *Personnel Psychology* between 1965 and 1982. Unlike the Hunter and Hunter results, these findings were not subjected to the numerous corrections advocated by Schmidt and Hunter (1977; 1990) other than corrections for sampling error. Schmitt et al. (1984) found cognitive ability tests were not superior to other predictors in terms of criterion-related validity as reported by Hunter and Hunter (1984). Overall average biodata validity was reported as nearly identical to that of general cognitive ability (.243 and .248, respectively) with biodata outperforming general cognitive ability in the prediction of performance ratings (.317 and .220, respectively), turnover (.209 and .141, respectively), and status change (.332 and .282, respectively). General cognitive ability yielded a mean validity of .437 compared to biodata mean validity of .226 in predicting achievement/grades.

Though meta-analytic efforts have examined simple criterion-related validities of various selection devices, very few primary research studies have directly compared the predictive validity of biodata against other selection techniques, such as general mental ability. A review of the biodata literature, a call on BIONET (a LISTSERV dedicated to biodata research), personal communication with leading biodata researchers, and an examination of all
criterion-related validity studies published in *Journal of Applied Psychology* and *Personnel Psychology* from 1964 through the present (from data previously reported in Schmitt et al., 1984, and Russell, Settoon, McGrath, Blanton, Kidwell, Lohrke, Scifres, & Danforth, 1994) yielded only seven studies reporting validity coefficients for both general cognitive ability and biodata predictors in the same sample with no studies reporting incremental validity of these predictors. Two of these seven studies labeled pure demographic information as biodata so it could be debated whether they should have been coded as biodata studies. Biodata has been found to add significant incremental validity when used with a battery of other selection devices (Reilly & Warech, 1990).

Given the lack of published primary research directly comparing predictors, analyses were performed that examined the criterion-related validity of biodata and general cognitive ability individually and when used in combination with each other. The following research questions were addressed:

**Research Question #4:** Is the biodata measure a valid predictor of performance?

**Research Question #5:** Is the general cognitive ability measure a valid predictor of performance?

**Research Question #6:** What are the relative contributions of biodata and general cognitive ability measures to performance prediction?
Research Question #7: Are 1) "g-loaded" and 2) "non-g-loaded" biodata items differentially related to the general cognitive ability measure?

Research Question #8: Do "g-loaded" biodata items, "non-g-loaded" biodata items, and the cognitive ability test exhibit incremental validity relative to one another?

On Biodata Adverse Impact

In addition to maximizing performance outcomes, corporations have both “push” and “pull” factors impinging on their selection devices. Organizations are “pushed” to comply with equal employment opportunity laws and regulations. For example, the EEOC Uniform Guidelines (1978) were put forth to help firms comply with Title VII of the 1964 Civil Rights Act. Organizations are “pulled,” by organizational performance needs and goals for a diverse workforce, which may bring a greater diversity of ideas. Selection research typically examines regulatory compliance of selection devices through adverse impact analysis. A selection device displays adverse impact if it results in “a substantially different rate of selection in hiring, promotion, or other employment decisions which works to the disadvantage of members of a race, sex, or ethnic group” (Gatewood & Feild, 1994, p. 104). This interest is also driven by the fact that general cognitive ability tests consistently report race differences of up to one standard deviation between majority and minority groups (US Employment Service, 1970). As workforce diversity continues to
increase, selection tests will need to be developed and implemented that predict job performance without having adverse impact on subgroups protected under federal legislation.

Hunter and Hunter (1984) suggested other predictors with less adverse impact than g may add to the validity of general cognitive ability measures. Several reviews found biodata is characterized not only by high predictive validity but also low adverse impact (Barge & Hough, 1986; Mitchell, 1994; Mumford & Stokes, 1992; Reilly & Chao, 1982; Reilly & Warech, 1990). Reilly and Chao (1982) evaluated criterion validity, adverse impact, and fairness of various predictors, finding biodata and peer evaluations had validities equal to standardized general cognitive ability tests. Reilly and Chao’s meta-analysis failed to find predictors having equal validity to standardized ability tests with less adverse impact.

The empirical nature of biodata has furthered the dustbowl empiricism stigma as mentioned earlier, but empirically derived biodata keys permit the virtual elimination of adverse impact. Very simply, once the key as been developed, response options demonstrating differential criterion prediction are dropped from the key. Pace and Schoenfeldt (1977) recommended that to maximize compliance under the EEOC Uniform Guidelines (1978) and for practical purposes as well, biodata item development be guided by job analysis with each item individually and the overall test score examined for adverse impact.
While adverse impact is not always determined by statistical analysis alone, the EEOC Uniform Guidelines (1978) and precedent established by the Supreme Court in Griggs v. Duke Power (1971) established the four-fifths rule as a statistical rule of thumb to alert possible violations of civil rights legislation. The four-fifths rule compares the selection ratio of the majority (the number of applicants selected versus the total number of majority applicants) to the selection ratios of each minority group (applicants selected/total number of applicants from the minority group). The ratio of any subgroup must be at least 4/5ths or 80% of the ratio of the majority subgroup for whom the device most favored (Gatewood & Feild, 1994) in order to "pass" the four fifths rule and not be seen as committing adverse impact.

Interestingly, in a review of the literature on subgroup differences in selection tests, Schmitt, Clause, and Pulakos (1996) found reporting of subgroup means diminished in the 1980s and early 1990s compared to studies conducted in the 1960s and 1970s. The decrease in reporting information on subgroup differences is troublesome given the need to ensure unbiased selection in a diverse workforce. It is not in the fields' best interests for researchers to put aside questions of adverse impact to focus solely on predictive validity issues.

Issues surrounding trade-offs between adverse impact and criterion-related validity have been recently highlighted in validation efforts for police officers in Nassau County, New York. Personality measures were given
greater weight than general cognitive ability measures to minimize adverse impact of selection tests against minorities (HR Strategies, 1995). This effort was criticized on the grounds that criterion validity was sacrificed to accomplish equal hiring rates among racial subgroups (Gottfredson, 1997). The Nassau County police validation effort resulted in the majority of weight being placed on personality measures reducing the selection system’s adverse impact (which had previously relied heavily on general cognitive ability measures).

In light of the decline in published data on adverse impact in personnel selection and recent controversies regarding use of alternative non-cognitive selection devices, a final goal of this study is to compare biodata and general cognitive ability selection devices for adverse impact. Specifically, the following research questions were addressed:

**Research Question #9:** When adverse impact response options are removed from the empirical key, is there a significant change in biodata criterion-related validity and adverse impact?

**Research Question #10:** Do general cognitive ability and biodata measures differ significantly in terms of adverse impact?

In sum, the literature on biodata continues to be plagued by the dustbowl empiricism label. Research is warranted to address this issue as biodata traditionally achieves high criterion validities and limited adverse impact compared to other selection devices. Research questions 1 - 3 addressed this relative absence of theory or more specifically, the absence of
biodata operationalizations being tightly linked to construct domains.

Biodata generates high average criterion validities, comparable to general cognitive ability measures. Very few primary research studies have been conducted which empirically examine predictive abilities of biodata and g. No research exists addressing possible overlap in construct domain between the two selection devices. Research questions 4 - 8 addressed the relative absence of direct comparisons of biodata and g, investigating the measures' construct domain overlap. Finally, the issue adverse impact will become more salient for organizations, especially given reliance on general cognitive ability measures for personnel selection. Research questions 9 and 10 addressed the urgent need for alternatives to g and its high adverse impact.
CHAPTER 3

METHODS

Sample

The data used in this study were obtained from a sample of candidates for the position for air traffic controller specialist (ATCS) from the period of October 1985 to January 1992. ATCS candidates must successfully complete a two stage selection process consisting of: 1) a written Office of Personnel Management (OPM) air traffic control selection test battery and 2) a nine-week training/screening program (the Screen).

The period of time in which data were collected was approximately 5 years after the strike of the Professional Air Traffic Controllers Organization (PATCO) in August 1981. This strike prompted the firing of all striking air traffic controllers by presidential order, resulting in a loss of the majority of the ATCS workforce. Post-strike ATCS trainee demographics changed meaningfully from those of pre-strike trainees. Collins, Nye, and Manning (1990) studied ATCS candidate demographics during pre-strike (1976-1981), immediate post-strike (1981-1983), and recent post-strike (1985-1987) time periods. Examination of three time periods was due to unprecedented hiring and training of over 8,000 replacement ATCS candidates in a 2-year period, a large amount of national strike publicity, a weak national job market, and highly publicized salaries of former ATCS which cumulatively resulted in attracting a different type of applicant to the job (Collins, Manning, & Taylor, 1984; Collins, Nye, & Manning, 1984).
Over two-thirds of pre-strike ATCS trainees had either prior aviation or air traffic control experience. More than two-thirds of post-strike applicants reported no prior aviation or air traffic control experience (Collins et al., 1990). The percentage of minority applicants also declined. Additional studies examining ATCS trainee demographic information pre- and post-strike report similar results (Taylor, VanDeventer, Collins, & Boone, 1983; VanDeventer, 1983a; VanDeventer, 1983b; VanDeventer and Baxter, 1984).

Data for the current investigation includes the recent post-strike era from 1985 to 1992. This study examined a sample of 11,405 ATCS candidates, of whom 5,814 completed the FAA Applicant Background Assessment biodata instrument (see explanation below). Criterion data (a training performance measure; see explanation below) were available on 10,114 candidates. The total sample was 82.3% male, 89% white, and the average age at time of entry into the profession was 25.9 years. Approximately 74% of the sample had no prior air traffic controller experience before applying for an ATCS position. Eleven percent had a high school degree, 55.6% had some college experience, 31.8% had a college degree, and 1.2% had earned an advanced degree prior to entry into the ATCS profession.

**Air traffic control specialist job specifications.** The ATCS job consisted of a complex set of tasks requiring high skill levels and use of cognitive abilities such as spatial perception, information processing, reasoning, and decision making (Della Rocco, Manning, & Wing, 1990). Harris (1986) reviewed...
previous air traffic controller (ATC) studies of abilities and psychological constructs to find effective predictors of ATC performance. She placed necessary abilities into three categories: 1) spatial perception; 2) verbal and non-verbal reasoning; and 3) mental manipulation of verbal or numeric concepts. Harris (1986) also found personality and temperament measures were not predictive of ATC performance. A more recent job analysis determined primary job attribute requirements of an ATCS to be perceptual speed, reaction time, memory, arithmetic reasoning, and spatial ability (Broach & Brecht-Clark, 1994).

The ATCS job encompasses three specialty options: en-route, terminal, and flight service station (FSS). En-route and terminal specialist positions both ensure separation of aircraft. En-route specialists monitor separation of aircraft traveling between airports and terminal specialists oversee separation of aircraft approaching or departing airports. Separation is accomplished through communications with pilots regarding altitudes and directions of flight. FSS specialists communicate with pilots on weather information, filing flight plans, and locating lost aircraft. FSS specialists have no aircraft separation responsibilities. FSS specialists require different knowledge, skills, and abilities than en-route and terminal specialists and thus require a unique selection program (Manning, Kegg, & Collins, 1988). This investigation examined candidates for en-route and terminal ATCS specialties.
Air Traffic Controller Specialist Selection Procedure and Measures

Candidates completed a multiple hurdle selection process over a minimum period of three years to become an ATCS. During this selection process candidates completed a number of paper and pencil tests and participated in training and job-related tasks that served as the data used in this research. Two primary hurdles included: 1) the OPM Air Traffic Controller Specialist Test Battery, followed by 2) the FAA Academy screening program (the Screen). Extensive reviews of the FAA ATC selection process can be found in Collins, Boone, and VanDeventer (1980) and Sells, Dailey, and Pickrel (1984).

Initial minimum requirements for consideration as an ATCS candidate included: high school education or equivalent, three years of general work experience (or college), 18 to 30 years of age, medical qualification, and a security clearance (Manning, Kegg, & Collins, 1988). The age requirement was mandated by Congress in 1972 and is exempt from the 1967 Age Discrimination in Employment Act as amended. Studies reported attrition for trainees 31 years of age or older to be two to three times higher than younger trainees (Collins, Boone, & VanDeventer, 1980). Manning et al. (1988) found supervisor job performance ratings of controllers in every age category over 40 were significantly below younger subgroups. VanDeventer and Baxter (1984) also reported a negative relationship between age and academy performance, providing post hoc job-related justification for the 1972 Congressional mandate.
Office of Personnel Management (OPM) test battery. The first stage of the selection process involved administration of the four hour ATCS written selection aptitude test battery by OPM. This battery contained the Multiplex Controller Aptitude Test (MCAT), the Abstract Reasoning Test (ART), and the Occupational Knowledge Test (OKT).

The Multiplex Controller Aptitude Test (MCAT) was a 110 item (86 minute) paper and pencil test designed to measure abilities required for air traffic control. The MCAT included traditional cognitive aptitudes found in many OPM tests such as arithmetic reasoning, data interpretation, table reading, and spatial relations (Manning, 1991; Manning, Kegg, & Collins, 1988)—job specifications which exist in the ATCS position (Harris, 1986). The test also contained job-related items such as identifying potential conflicts between aircraft in simulated traffic on air route maps (Dailey & Pickrel, 1984).

Test-retest reliability of the MCAT was estimated at .60 and parallel forms reliability ranged from .42 to .89 in a sample of 617 newly hired controllers (Rock, Dailey, Ozur, Boone, & Pickrel, 1981). Available data suggested the MCAT had acceptable reliability but was vulnerable to practice effects (Broach, 1997). Only first time applicants were included in the current sample to control for practice effects. The MCAT was found to be significantly correlated with both Screen performance and on-the-job training performance across many cohorts of ATCS candidates. The criterion validity of the MCAT in predicting Screen performance ranged from .24 (corrected correlation; $r_c =$
.48) to .28 (r_c = .55). All reported corrected correlations are corrected for range restriction on the predictor. For additional MCAT criterion validity evidence see Boone (1979); Manning, Della Rocco, and Bryant (1989); Mies, Colmen, and Domenec (1977); Schroeder, Dollar, and Nye (1990).

The Abstract Reasoning Test (ART) was a timed (35 minute) 50-item paper and pencil test of ability to infer relationships between symbols. Items included on the test involved letter series and figure classification. The following is an example letter series item:

```
ARCSETG_ _  a. HI  b. HU  c. UJ  d. UI  e. IV
```

The ART incrementally contributed to prediction of the Screen simulation score of 1827 ATCS students (Boone, 1979). Schroeder et al. (1990) reported significant correlations between the ART and Screen success (pass/fail) .12 (r_c = .45) and the Screen composite .17 (r_c = .26). Broach and Manning (1994) suggested the ART was predictive of both ATCS Screen and on-the-job performance.

The MCAT and ART scores were weighted 2 and 1 (respectively), combined, and standardized with a mean of 70 and a maximum of 100 resulting in a composite aptitude score (APT). Numerous FAA studies examined the criterion-related validity of the APT with various criteria. For example, Manning et al. (1989) found the APT significantly correlated with Screen performance .31 (r_c = .61) in a sample of en route ACTSs. Schroeder et al., (1990) found the APT significantly correlated with Screen performance (pass/fail) .17 (r_c = .46)
and with Screen composite scores .21 (rc = .54). Under current policy, candidates must score at least 70 on the APT to be considered for the second stage of selection, attendance at the FAA Academy Screen.

Two additional measures—the Occupational Knowledge Test (OKT) and Veterans Preference Credit (VET)—allowed "extra credit points" to be added to a minimum score of 70 on the APT. The OKT was a 80 item, 50 minute test designed to assess and assign extra credit points to candidates demonstrating ATC job knowledge (Dailey & Pickrel 1984). Candidates had the opportunity to earn up to 15 extra credit points by correctly answering OKT items (0-51 items correct = 0 extra points; 52-55 items correct = 3 extra points; 60-63 items correct = 10 extra points; 64-80 items correct = 15 extra points). Dimensions of ATC occupational knowledge on the OKT included: air traffic rules, airport traffic procedures, in-flight traffic control procedures, communications operating procedures, flight assistance service procedures, air navigation and aids to navigation, and aviation weather.

Veterans preference credit (VET) awarded candidates with prior military experience extra credit points toward their overall OPM test battery score. Veterans received 5 extra points on their OPM test score or 10 extra points if the veteran had a disability related to their military status. Veterans also had priority status in hiring (Aul, 1997). Extra points earned through demonstrated job knowledge (OKT) or previous military experience (VET) could not help candidates pass the aptitude screening stage (i.e., achieve the minimum score
of 70), but it could increase an applicant’s rank order position on the list from which actual selection was made (Manning, 1991). Regardless, the selection process was quite competitive. While 70 was the minimum score needed for qualification, typically only candidates with OPM ratings of 90 or above were selected (Dr. Dana Broach, personal communication, July 1997; Manning et al., 1990).

The composite predictor used by the FAA to select candidates into the training Screen was the overall OPM rating (RAT). All previously described predictors (MCAT, ART, OKT, and VET) were summed to form this overall rating. This rating was used to rank eligible candidates on a register from which selections to the training Screen were made. The following two equations summarize the calculations performed to obtain the APT and RAT scores:

\[
APT = 2MCAT + ART
\]
\[
RAT = APT + (OKT + VET)
\]

For the current investigation, only the APT was analyzed as this is the closest representation of ATCS applicant general cognitive ability. The OKT and VET are not measures of general cognitive ability and were excluded. Interestingly, prior to 1964 the screening and selection of ATC candidates involved no formal assessment of applicant mental abilities or aptitudes (Cobb & Matthews, 1972).
Some ATCS applicants completed biodata inventories as part of a research project undertaken to develop new procedures for possible future use in competitive examinations for ATCS selection. This biodata information was collected on candidates during the selection process but was not used in selection decisions. The biodata measure to be examined is the FAA Applicant Background Assessment.

The Applicant Background Assessment is an 142 item biodata questionnaire. The ABA was developed based on: 1) a review of qualification standards for ATCS, 2) a review of job analyses conducted by the FAA, 3) a review of previous biodata work done at the FAA, 4) interviews with training staff members to determine characteristics of ATCSs that differentiate those who perform better in training and those that fail training, and 5) interviews with ATCS supervisors to ascertain characteristics differentiating good and poor ATCSs. The items included on the ABA were limited to those dealing with experiences under applicant control (Dr. Dana Broach, personal correspondence, October, 1997). No construct or criterion-related validity information was available on the ABA.

FAA Academy Screen. The second stage of the selection process was the FAA ATCS Nonradar Screen, a nine-week initial training program administered by the FAA Academy in Oklahoma City, OK. The Screen composite score was the primary criterion measure used in this study. The Screen taught candidates with no knowledge of air traffic control enough about
the job to assess potential advancement to full performance level as an operating ATCS (Della Rocco, Manning, & Wing, 1990). The Screen provided candidates with knowledge of basic air traffic rules and procedures then tested applicant knowledge through exams and laboratory simulations.

Three categories of performance assessments were included in the overall screen composite score (Screen): 1) paper and pencil exams, 2) simulations, and 3) final examination. These categories were weighted 20%, 60%, and 20%, respectively, and summed to form an overall Screen composite score. The first group of assessments included a series of multiple-choice tests assessing candidate ability to acquire and retain basic job knowledge. The second set of assessments included systematic evaluations of trainee performance on six 30 minute laboratory simulations of non-radar air traffic control. The simulations were scored using 1) an average instructor technical assessment of number of errors, 2) an average instructor assessment of trainee performance, and 3) the average 5 out of 6 highest scores on the individual laboratory simulations. The final portion of the Screen was a multiple choice final exam assessing trainees ability to apply ATC rules and procedures. Trainees must have scored at least 70 out of a possible 100 on the composite to pass the Screen and be eligible for on-the-job training (Aul, 1991; 1997; Della Rocco, et al., 1990; Young, Broach, & Farmer, 1996).

Early FAA studies found laboratory simulation portions of the Screen to be the most accurate predictors of ATCS success on-the-job (Cobb, 1962,
1965; Trites, 1961, 1965) and provided justification for their contribution of 60% to the Screen composite score. Recent criterion-related validity studies provided support for the overall Screen composite. Specifically, Manning et al., (1989) found the SCREEN significantly predicted attrition, supervisor ratings, and field training status (i.e., whether one completed on-the-job training, was still training on-the-job, switched options, or failed). Della Rocco et al. (1990) examined a cohort of Screen graduates assigned to the en route option and found a significant correlation between Screen composite score and field training status (r = -.24, r_c = -.44). All results examining Screen composite correlations were attenuated due to range restriction on the Screen.

Broach and Manning (1994) investigated the Screen's ability to predict subsequent performance in on-the-job radar training after 1 to 2 years as en route and terminal ATCSs. Screen performance was significantly correlated with on-the-job en route radar training performance .28 and on-the-job terminal radar training performance .31. After correcting for range restriction due to explicit selection on the Screen composite score, both correlations increased to .50. Broach and Manning (1994) also found the Screen composite added incremental validity over aptitude ratings accounting for an additional 8% of variance in on-the-job en route and 10% of variance in terminal training performance. After correcting for range restriction, the incremental variance explained by Screen performance was 20% and 16%, respectively. In sum,
numerous studies suggested performance on the Screen was highly predictive of on-the-job ATCS performance.

The Screen was developed in response to a US Congressional House Committee on Government Operations recommendation to "provide early and continued screening to insure the prompt elimination of unsuccessful trainees and relieve the regional facilities of much of this burden" (US Congress, 1976, p. 13). Prior to implementation of the Screen in 1976, training attrition occurred on average 2 to 3 years into an individual's tenure, resulting in high turnover costs (Cobb, Mathers, & Nelson, 1972; Manning, 1991). Prior to Screen implementation, the field training attrition rate was 41%. After Screen implementation, field attrition dropped to 8% with most of attrition occurring during the 9-week training Screen (Della Rocco et al., 1990). Aul (1991) estimated approximately 40% of participants historically failed the Screen and were terminated.

From October 1985 to January 1992 less than 10% of over 206,000 candidates who took the OPM test were selected to advance to the Screen. Of 12,869 candidates who advanced to the Screen, 7,091 successfully passed and were assigned to an air traffic control facility for on-the-job training (Broach & Brecht-Clark, 1994). While training on-the-job, the ATCS is essentially an "apprentice" ATCS who works under direct supervision of a senior ATCS. Field training was conducted on an "up or out" basis (Aul, 1991), i.e., apprentice ATCSs had to progress toward full performance level (FPL) air traffic controller
status or be terminated. Upon successful completion of on-the-job training, the ATCS earns the title of FPL ATCS. The system was designed to place more successful trainees in terminal controller positions at high traffic airports, but actual practice resulted in supply and demand dictating trainee placement in terms of location and ATCS type (personal communication, Dr. Dana Broach, June 1997).

The Screen performance composite was the FAA’s primary performance criterion. Adequate job performance measures were not available for this sample. This is due in part because: 1) there was no formal performance appraisal process for ATCSs (although a formal performance appraisal process is currently being developed under the direction of Dr. Walter Borman; Dr. Dana Broach, personal communication, June 1997), 2) job performance variability was minimal given the critical public safety nature of the job, and 3) existing job performance measures were not precise due to union agreements that mandate controllers be evaluated on a dichotomous, satisfactory/non-satisfactory criterion.

To recap, the predictor and criterion measures used in this study were made available from FAA archival data on ATCS candidates from years 1985 to 1992. The predictors examined include an experimental biodata instrument and a composite general cognitive ability measure. The criterion for this study was the Screen composite score. Permission to use these FAA data is found in the Appendix.
Analyses

Analyses are described in order of research question addressed. Three sets of research questions addressed: 1) construct validity of items anchored in biodata theory, 2) individual and incremental construct validity of biodata and general cognitive ability measures, and 3) adverse impact of biodata versus general cognitive ability. Examination of whether biodata items demonstrate construct validity involved empirically examining predictions derived from construct domains drawn from Mumford, Stokes, and Owens' (1990) ecology model. Relationships between subjects' responses to these items and a FAA performance measure was examined for consistency with ecology model predictions in a sample of ATCSs. Analyses estimated the degree to which biodata and general cognitive ability tests individually and incrementally predict performance. Finally, the degree of adverse impact was estimated for individual predictors and predictor combinations.

Empirical keying was used to score the biodata instrument examined in this study. A wide variety of empirical keying methods exist, but prior research suggests methods directly estimating strength of relationships between biodata response options and criterion do best. (Devlin, Abrahams, & Edwards, 1992). The point biserial correlation ($r_{pb}$) between each response option and the criterion were used as weights for the empirical key. The point biserial correlation is a special case of the Pearson product-moment correlation ($r$) applicable when correlating a truly dichotomous (e.g., a response option that is
chosen or not chosen) with a continuous variable (e.g., a performance measure). In this situation, the Pearson product moment correlation formula reduces to the more simplified formula for $r_{pb}$. The point biserial is a more efficient estimate of the strength of this necessarily linear relationship (Nunnally & Birnberg, 1995) in that it uses all observations in a sample (i.e., it does not throw out the middle one-third of performers as is often done in the construction of empirical biodata keys; Mumford & Owens, 1987).

**Issue I: Biodata Construct Validity**

The first set of analyses sought to answer the general question: Do biodata items tap construct domains identified by biodata theory? In order to test this research question a number of analyses and procedures were undertaken including: 1) sorting items into theory-based construct domains, 2) testing for construct validity of sorted domains through convergent and discriminant validity analyses and confirmatory factor analyses, and 3) examining criterion-related validity. Items were Q-sorted into construct domains drawn from the respective models and subjected to confirmatory factor analyses. An empirical key was developed (described below) for biodata items using point biserial correlations to evaluate biodata criterion validity. This analysis was conducted on the sample of all ATCSs who completed the biodata questionnaire and on whom criterion measures where obtained. Biodata scale scores were correlated with criterion measures. The following research questions addressed if biodata items tap theory-based construct domains:
Research Question #1: Do items exhibit factor loadings in a manner consistent with a priori theory-based construct domains?

Research Question #2: Do relationships among biodata scale scores derived for latent biodata constructs yield convergent and discriminant validities consistent with theory-based construct domains?

Research Question #3: Do biodata items sorted into theory-based construct domains demonstrate higher criterion-related validity than non-theory-based items?

The following analyses addressed the first set of research questions:

Step 1: Q-sort. Biodata items from the Federal Aviation Administration (FAA) Applicant Background Assessment (ABA) were subjected to a Q-sort procedure by 5 judges with knowledge of biodata applications (1 PhD and 4 advanced doctoral students in human resource management related fields). Judges were asked to sort biodata items into biodata theory construct domains hypothesized to predict future performance criteria in the ecology model. The ecology model posited social resources, personality resources, intellectual resources, filtering processes, and choice processes as construct categories influencing environmental outcomes (e.g., performance).

Q-sorters read descriptions of each construct domain and were asked to place items into construct domains that, in their opinion, item content most
represented. Sorters were also asked to place those items not sorted into any construct domain into a separate pile. A pilot Q-sort was conducted to determine if descriptions were clear and gather information on the ease with which items fit into piles. Pilot sorters were also asked to 1) note which items (if any) fit in multiple constructs and 2) a priori see if they could identify additional constructs underlying items that did not fall into a priori construct domains. This was done initially with the pilot group to fine-tune instrument clarity.

Some items were judged in the pilot test as potentially multi-dimensional. In light of this, subsequent Q-sorters were asked to sort each item into a primary domain and, if necessary, list secondary domains as well. This was done to minimize procedure difficulty. No additional construct domains were gleaned from the Q-sort procedure for this set of items. Consensus discussion among sorters took place to resolve disagreements in item classification.

**Step 2: Assessment of construct validity.** Confirmatory factor analyses (CFA) were conducted to determine whether each group of items emerging from the Q-sort yield factor loadings consistent with the a priori construct domains. Convergent and discriminant validity among the constructs were examined for consistency with theory-based relationships. If CFA did not provide sufficient evidence that Q-sorted item groupings tapped a priori theory-based constructs, exploratory factor analysis (EFA) was used to determine if other possible item groupings exist.
Step 3: Determination of criterion-related validity. If construct validity is suggested for all or some theory-based constructs, criterion-related validities was assessed for construct-specific scales. Additionally, theory-based items and items that could not be explained by theory were examined separately. An empirical key was developed from the point biserial correlations between each response option and criterion within the key development sample for each biodata sub-scale examined. Correlations were used as response option weights in all empirical keys. Biodata scale scores were set equal to sum of the correlations associated with the response options each individual selected.

Issue II: Biodata Criterion Validity

The second set of analyses addressed the general question: What is the relative contribution of biodata and general cognitive ability measures to performance prediction? Before addressing relative contributions of biodata and g, criterion validity of both biodata and general cognitive ability measures was examined individually. Each measure was correlated with the performance criterion to determine simple criterion validities. Additional analyses addressed whether 1) biodata items found to tap general cognitive ability construct domain (i.e., g-loaded items), as evidenced by results of the Q-sort, CFA, or EFA, show incremental validity over the general cognitive ability measure and 2) items that do not tap intelligence (i.e., non-g loaded items) show incremental validity over the general cognitive ability measure and g-loaded biodata scales.
The following research questions addressed biodata and general cognitive ability measure criterion validity:

**Research Question #4:** Is the biodata measure a valid predictor of performance?

**Research Question #5:** Is the general cognitive ability measure a valid predictor of performance?

**Research Question #6:** What are the relative contributions of biodata and general cognitive ability measures to performance prediction?

**Research Question #7:** Are 1) "g-loaded" and 2) "non-g-loaded" biodata items differentially related to the general cognitive ability measure?

**Research Question #8:** Do "g-loaded" biodata items, "non-g-loaded" biodata items, and the cognitive ability test exhibit incremental validity relative to one another?

The following analyses were conducted to address this research question:

**Step 1: Determine individual criterion-related validity.** Individual criterion related validity was determined by empirically keying the entire set of biodata items and estimating the resultant biodata score's correlation with the criterion. The g measure was also correlated with the criterion.
Step 2: Derive $g$ and non-$g$ loaded biodata items. Items Q-sorted into the general cognitive ability ($g$) construct domain with supportive construct validity evidence from CFA and/or EFA results were extracted and examined for criterion validity separately from other biodata items (i.e., "non-$g$-loaded" items).

Step 3: Determine incremental criterion validity. Incremental criterion validity was determined using hierarchical regression analyses. The criterion was regressed onto the following predictors: 1) non-$g$ biodata items, 2) $g$ biodata items, 3) APT, and 4) all biodata items. A second set of equations regressed the APT and the three sets of biodata items (all biodata items, $g$ items, and non-$g$ items) onto the criterion. These equations are summarized in Table 3.1.

Table 3.1
Hierarchical Multiple Regression Analysis Equations

<table>
<thead>
<tr>
<th>Equation</th>
<th>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{biodata non-g items}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation 2</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{biodata g items}}$</td>
</tr>
<tr>
<td>Equation 3</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{all biodata items}}$</td>
</tr>
<tr>
<td>Equation 4</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{cognitive ability test}}$</td>
</tr>
<tr>
<td>Equation 5</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{cognitive ability test}} + \beta_2 X_{\text{all biodata items}}$</td>
</tr>
<tr>
<td>Equation 6</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{cognitive ability test}} + \beta_2 X_{\text{biodata g items}}$</td>
</tr>
<tr>
<td>Equation 7</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{cognitive ability test}} + \beta_2 X_{\text{biodata non-g items}}$</td>
</tr>
<tr>
<td>Equation 8</td>
<td>$Y_{\text{criterion}} = \beta_0 + \beta_1 X_{\text{cognitive ability test}} + \beta_2 X_{\text{biodata g items}} + \beta_3 X_{\text{biodata non-g items}}$</td>
</tr>
</tbody>
</table>
In sum, evidence suggesting biodata added independent and non-overlapping predictive power to g was forthcoming to the extent that 1) CFA confirmed g and non-g biodata items, 2) g biodata items were highly correlated with the cognitive ability measure and 3) non-g biodata items incrementally added to the criterion validity obtained by the general cognitive ability measure.

**Issue III: Biodata Adverse Impact**

The third set of analyses sought to answer questions relating to the degree of adverse impact on racial subgroups in biodata and general cognitive ability measures. The final research questions of this study addressed the following adverse impact issues:

**Research Question #9:** When adverse impact response options are removed from the empirical key, is there a significant change in biodata criterion-related validity and adverse impact?

**Research Question #10:** Do general cognitive ability and biodata measures differ significantly in terms of adverse impact?

The following analyses were performed to address these research questions:

**Step 1: Biodata response option-level adverse impact analysis.** Each response option was examined for compliance with the four-fifths rule (*Griggs v. Duke Power*, 1971; *Uniform Guidelines*, 1978) to determine if blacks and whites answered with differential frequency (i.e., one group selecting a response option at 80% or less the other group’s rate).
Step 2: Development of biodata sub-scale with only non-adverse impact response options. Two empirical keys were developed from the key development sample. One key contained all 710 biodata response options, while a second key included only those response options that passed the four-fifths criterion.

Step 3: Determination of standardized mean difference biodata scores. The standardized mean difference between blacks and whites was calculated for the biodata scale scores with and without adverse impact response options and performance on the general cognitive ability test.

Step 4: Determination of biodata scales criterion validity. The biodata scales with and without racial subgroup adverse impact response options were also examined to determine if there was a significant decrement in criterion-related validity when response options demonstrating adverse impact were removed from the empirical key used to score the biodata instrument. A Hotelling-Williams test was performed to determine whether criterion validities for the two biodata scales were significantly different. The Hotelling-Williams test allows comparison of two correlations that are dependent on each other (Bobko, 1995). These correlations were necessarily dependent due to computation on the same subject sample (the cross-validation sample), the use of a common variable (training performance), and the use of common predictors (response options with no adverse impact entered both keys).
Step 5: Analysis of test fairness. The biodata scales both with and without adverse impact response options and the general cognitive ability test were examined for violations the Cleary (1968) model of test fairness (referred to as the "regression model" by the EEOC Uniform Guidelines, 1978). A selection device is "fair" under the Cleary model if the regression coefficient, $b_3$, fails to reject $H_0: b_3 = 0$ in the equation below:

$$Y_{predicted} = b_0 + b_1X_{selection~device} + b_2X_{race} + b_3X_{selection~device} \times X_{race} + e$$

Step 6: Comparison of adverse impact rates in the sample. A final set of analyses compared adverse impact that biodata and the cognitive ability test might have in this particular data set. Cut scores at the 20th, 40th, 60th, and 80th percentiles of the sample predictor distributions were used as cut scores to illustrate the degree of adverse impact for each instrument. The selection rate for blacks and whites for each cut score on the biodata predictor scales and cognitive ability measure were then subjected to the four-fifths rule to determine if adverse impact existed.
CHAPTER 4
RESULTS

This chapter reports results from analyses addressing the ten questions posed in this research. A Q-sort procedure, response-option based empirical key, and confirmatory and exploratory factor analyses were used to address research questions 1 through 3. These research questions addressed the degree to which biodata items captured latent constructs described in the ecology model. Simple correlational and hierarchical multiple regression analyses were used to address research questions 4 through 8, which addressed the relative contributions to prediction of biodata and cognitive ability. Research questions 9 and 10 addressed the degree of racial subgroup adverse impact in biodata scales and a general cognitive ability measure. Adverse impact was examined at the response option level for impact on standardized mean subgroup scores and on criterion-related validity. Moderated multiple regression analysis was used to assess the test fairness of the biodata instrument (including and excluding adverse impact response options) and the general cognitive ability test. Additional analyses were performed to determine effects of various cut scores on general cognitive ability test and biodata inventory adverse impact.

Means and Correlations

Descriptive statistics are presented in Table 4.1, while Table 4.2 reports intercorrelations and cross-validities among biodata, cognitive ability, and
Table 4.1
Descriptive Statistics for the Entire, Key Development, and Cross-Validation Samples

<table>
<thead>
<tr>
<th></th>
<th>Entire Sample</th>
<th>Key Development Subsample</th>
<th>Cross-Validation Subsample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SDx</td>
</tr>
<tr>
<td>Screen</td>
<td>10014</td>
<td>71.66</td>
<td>11.35</td>
</tr>
<tr>
<td>Cognitive Ability Test Biodata</td>
<td>10869</td>
<td>91.46</td>
<td>5.02</td>
</tr>
<tr>
<td>Interpersonal Personality</td>
<td>6036</td>
<td>1.52</td>
<td>1.07</td>
</tr>
<tr>
<td>Cognitive Ability Motivation</td>
<td>5681</td>
<td>1.036</td>
<td>.215</td>
</tr>
<tr>
<td>Self-Perception Non-theory Items</td>
<td>5766</td>
<td>1.020</td>
<td>.103</td>
</tr>
<tr>
<td></td>
<td>5567</td>
<td>1.081</td>
<td>.868</td>
</tr>
<tr>
<td></td>
<td>5141</td>
<td>1.264</td>
<td>.285</td>
</tr>
<tr>
<td></td>
<td>5810</td>
<td>1.027</td>
<td>.069</td>
</tr>
<tr>
<td></td>
<td>5706</td>
<td>1.077</td>
<td>.087</td>
</tr>
<tr>
<td></td>
<td>4961</td>
<td>1.429</td>
<td>.422</td>
</tr>
<tr>
<td></td>
<td>4779</td>
<td>1.215</td>
<td>.731</td>
</tr>
</tbody>
</table>

performance criterion measures. The biodata inventory was additionally examined by dividing it into sub-scales based on: 1) items Q-sorted according to ecology model constructs, 2) non-theory based items, 3) five groupings of randomly chosen theory based items, 4) non-g items (all items not classified as tapping cognitive ability), and 5) response options that did not adversely impact either minority or majority group members. Biodata criterion validities,
reported in Table 4.2, were obtained from a cross-validation sample using empirical keys derived from a key development sample to optimally predict performance on the Screen. The entire sample was randomly divided into key development (80%) and cross-validation (20%) samples for purposes of empirical keying. It was necessary to cross-validate biodata criterion validities to reduce the possibility that the key capitalized on chance relationships in the data set. Sample sizes between variables differ because some subjects did not complete a biodata instrument, had missing data on the biodata instrument, or did not have a criterion measure.

Unadjusted cross-validities range from .365 for empirically keyed response options taken from the entire inventory (.440 adjusted for indirect range restriction due to selection on the FAA selection battery) to .065 for empirically keyed response options taken from the four items Q-sorters categorized as Self Perception items. Response options from items Q-sorted onto Cognitive Ability and Motivation ecology model construct domains yielded the highest sub-scale cross-validities (.296 and .209, respectively). Curiously, three of six cross-validities reported in Table 4.2 column 1 are larger than validities observed in the key development sample. Typically, the cross-validity is less than the correlation obtained from the key development sample. This uncommon but not impossible finding speaks to the reliability and generalizability of the empirical keys. The cross-validities obtained from the biodata instrument and sub-scales provide criterion-related validity evidence.
Table 4.2
Intercorrelations for Entire Sample, Key Development Sample, and Cross-Validation Subsample

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Screen</td>
<td>-</td>
<td>.184</td>
<td>.192</td>
<td>.155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cognitive Ability Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Entire Biodata Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Interpersonal Biodata Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Personality Biodata Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Cognitive Ability Biodata Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Motivation Biodata Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Self-perception Biodata Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Non-g Biodata Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Non-Adverse Impact Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The three values reported in each cell correspond with correlations taken from the entire sample (upper left cell), key development sample (middle cell), and the cross-validation sample (lower right cell), respectively.
Biodata Construct Validity

Regarding biodata construct validity, the following research questions were posed:

**Research Question #1**: Do items display psychometric characteristics (e.g., content validity, item factor analytic loadings, internal consistency reliability) consistent with theory-based construct domains?

**Research Question #2**: Do relationships among scale scores derived for latent biodata constructs yield convergent and discriminant validities consistent with theory-based construct domains?

**Research Question #3**: Do items sorted into theory-based construct domains demonstrate higher criterion-related validity than non-theory-based items?

Addressing Research Question 1, the Q-sort procedure used to group biodata items into the ecology model construct domains served as an assessment of item content validity. Content validity reflects the degree to which items are representative of construct domains being sampling from and is assessed via expert judgment. The five Q-sort judges agreed on 116 of the 142 items (82% agreement). Consensus was reached via group discussion on 26 remaining items. Judges agreed 11 of the 142 items could not be placed in any of the construct domains (these were subsequently categorized as "non-theory" based items). Hence, a total of 131 items were categorized into the
ecology model construct domains. High inter-rater agreement provides
evidence for content validity for these item groupings.

Confirmatory factor analyses (CFA) using LISREL 8.2 (Joreskog &
Sorbom, 1996) assessed congruence of latent factor structure with a priori item
groupings. Fifteen biodata items with categorical response options were
excluded from all CFA analyses due to fact that the scales scores on
categorical responses are meaningless and uninterpretable in factor analyses.
CFA performed at the item level can be inconclusive due to large numbers of
parameters to be estimated and possible violations of multivariate normality
assumptions (March, Antill, & Cunningham, 1989; Russell, Kahn, Spoth, &
Altmaier, 1998; West, Finch, & Curran, 1995). Consequently, common CFA
practice involves using averages of item groupings or "parcels" as indicators
(Schau, Stevens, Dauphine, & Del Vecchio, 1995).

Common factor analyses with oblique factor rotation within Q-sort
categories were performed on biodata items within Q-sort categories to identify
internally consistent, unidimensional parcels (Drasgow & Kanfer, 1985; Kishton
& Widaman, 1994). EFA was used to guide item parcel construction to ensure
each parcel represented only one underlying factor as per Drasgow and
Kanfer's (1985) recommendations. This analysis was done to avoid
constructing parcels with two or more underlying factors, a major disadvantage
of parcel construction (West, Finch, & Curran, 1995).
In some instances factor analyses yielded multiple factors within each Q-sort group (see Table 4.3). Interpretable factors guided development of the item parcels for each construct category. Factor analyses performed on Q-sorted interpersonal biodata items analyses suggested two meaningful factors: 1) superiors’ and peers’ views of applicant’s interpersonal skill and 2) high school and college related interpersonal skill. The personality item factor analyses yielded two factors: 1) superiors’ views of applicant personality and 2) peers’ views of applicant personality.

Exploratory analyses conducted on g-loaded biodata items yielded three interpretable underlying factors: 1) evidence of cognitive ability in college, 2) prior superiors’ views of applicant general cognitive ability, and 3) evidence of general cognitive ability in high school. Two interpretable factors emerged from Q-sorted motivation items: 1) job-related motivation, and 2) school-related motivation. The Q-sort yielded only two items in the self-perception category. Each was treated as a separate indicator.

Loadings among factors within each ecology construct category were clean. Average factor loading on dominant factors was .58, while average factor loading on non-dominant factors was .05. The alpha levels for each factor were also relatively high (see Table 4.3) given the items were Q-sorted into the construct domains post-hoc and not developed a priori with these domains in mind. In sum, a total of 11 parcels were created representing the 5
ecology model constructs: two interpersonal, two personality, three cognitive ability, two motivation, and two self-perception indicators.

Item parcels were loaded onto the ecology model constructs in which their respective items were Q-sorted in a confirmatory factor analysis. Internal consistency reliabilities for parcels were much higher than for initial Q-sort groupings (see Table 4.3), suggesting that using parcels as indicators was more meaningful than grouping all Q-sorted items into one indicator. A number of goodness of fit indices were examined. Chi-square values, goodness of fit index (GFI; Joreskog & Sorbom, 1989), adjusted GFI, normed fit index (NFI; Bentler & Bonett, 1980), parsimonious normed fit index (PNFI; Mulaik, James, Alstine, Bennett, Lind, & Stilwell, 1989), comparative fit index (CFI; Bentler, 1990), and root mean square error of approximation (RMSEA) were used to assess model fit. Each goodness of fit index takes into account different aspects of model fit and range from zero to 1.000. Higher values suggest greater model fit, with the exception of RMSEA, where values less than .05 are good and values as high as .08 are reasonable (Browne & Cudeck, 1993). A chi-square equal to its degrees of freedom represents perfect fit, while a large number indicates lack of fit.

The GFI favors models with many estimated parameters, while the CFI favors more parsimonious models. The NFI reflects the proportion of total information accounted for by a model, the PNFI takes into account both the goodness of fit and parsimony of the model, and the RMSEA takes into
Table 4.3
Exploratory Factor Analysis Results within Q-sort Ecology Model Construct Categories†

<table>
<thead>
<tr>
<th>Interpersonal Items</th>
<th>Loadings*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α = .74</strong></td>
<td>1</td>
</tr>
<tr>
<td>My previous supervisor (or teachers if not previously employed) would rate my oral communication skills as...</td>
<td>.693</td>
</tr>
<tr>
<td>My peers would rate my interpersonal skills as...</td>
<td>.626</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my supervisory potential as...</td>
<td>.602</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my ability to get along with others as...</td>
<td>.581</td>
</tr>
<tr>
<td>My peers would rate my skill in influencing people to my point of view as...</td>
<td>.563</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my responsiveness to others' viewpoints</td>
<td>.536</td>
</tr>
<tr>
<td>My previous supervisor (or teacher...) would rate my skill at speaking before a group as...</td>
<td>.511</td>
</tr>
<tr>
<td>Which of the following would your peers say describes your behavior in a social situation?</td>
<td>.507</td>
</tr>
<tr>
<td>Compared to my peers, I find myself leading others...</td>
<td>.486</td>
</tr>
<tr>
<td>Which of the following would your peers say best describes your behavior in a group situation?</td>
<td>.450</td>
</tr>
<tr>
<td>Compared to my co-workers, people come to me for advice...</td>
<td>.432</td>
</tr>
<tr>
<td>High school classmates would most likely describe my leadership in extracurricular activities as...</td>
<td>.201</td>
</tr>
<tr>
<td>High school classmates would most likely describe my participation in extracurricular activities as...</td>
<td>.143</td>
</tr>
<tr>
<td>Number of elected offices in high school...</td>
<td>.023</td>
</tr>
<tr>
<td>Number of college clubs and organized activities in which I participated</td>
<td>-.100</td>
</tr>
<tr>
<td>Number of student office to which I was elected in college...</td>
<td>-.094</td>
</tr>
<tr>
<td>In organizations to which I belong, my participation is best described as...</td>
<td>.044</td>
</tr>
<tr>
<td>Relative to other high school students, my classmates would most likely describe my leadership skills as...</td>
<td>.397</td>
</tr>
<tr>
<td>Relative to other high school students, my classmates would most likely describe my interpersonal skills as...</td>
<td>.399</td>
</tr>
<tr>
<td>Number of elected offices (other than HS or college) I have held in the past 5 years</td>
<td>-.003</td>
</tr>
<tr>
<td>The number of years of leadership experience I have had (such as work supervisor, scout patrol leader, school or social club president, athletic captain, etc.) is...</td>
<td>.287</td>
</tr>
</tbody>
</table>

**α*** .81 .68
(Table 4.3 continued)

<table>
<thead>
<tr>
<th>Personality Items</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
</tr>
<tr>
<td>α = .27</td>
<td></td>
</tr>
<tr>
<td>87 My previous supervisor (or teachers if not previously employed) would rate my</td>
<td>.543</td>
</tr>
<tr>
<td>dependability as.</td>
<td></td>
</tr>
<tr>
<td>96 My previous supervisor (or teachers...) would rate my self-control as...</td>
<td>.579</td>
</tr>
<tr>
<td>108 My previous supervisor (or teachers...) would rate my attention to detail...</td>
<td>.538</td>
</tr>
<tr>
<td>7 My high school teachers would describe my self-discipline as...</td>
<td>.411</td>
</tr>
<tr>
<td>119 My peers would describe my aggressiveness as...</td>
<td>-.152</td>
</tr>
<tr>
<td>118 My peers would describe me as a person who takes chances...</td>
<td>-.080</td>
</tr>
<tr>
<td>125 My peers would describe my self-confidence as...</td>
<td>.339</td>
</tr>
<tr>
<td>15 During my years in high school, I was singled out for disciplinary reasons:</td>
<td>.263</td>
</tr>
<tr>
<td>57 Prior to accepting my first job in my present job series, I have been employed</td>
<td>.001</td>
</tr>
<tr>
<td>that of my present job for...</td>
<td></td>
</tr>
<tr>
<td>124 My peers would probably say that having someone criticize my bothers me:</td>
<td>.220</td>
</tr>
<tr>
<td>α = .58</td>
<td>.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Cognitive Ability Items</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>α = .67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 Class standing in college.</td>
<td>.851</td>
<td>.040</td>
<td>-.032</td>
</tr>
<tr>
<td>31 College grade received most often.</td>
<td>.845</td>
<td>.023</td>
<td>-.043</td>
</tr>
<tr>
<td>35 Overall college GPA.</td>
<td>.898</td>
<td>.009</td>
<td>-.050</td>
</tr>
<tr>
<td>34 GPA in college major.</td>
<td>.801</td>
<td>.013</td>
<td>-.052</td>
</tr>
<tr>
<td>25 Number of times you made the Dean's List in college.</td>
<td>.770</td>
<td>-.010</td>
<td>.010</td>
</tr>
<tr>
<td>32 First 2 years college GPA</td>
<td>.745</td>
<td>-.002</td>
<td>-.021</td>
</tr>
<tr>
<td>40 College English grade received most often.</td>
<td>.725</td>
<td>.026</td>
<td>.037</td>
</tr>
<tr>
<td>33 GPA after first two years of college.</td>
<td>.724</td>
<td>.007</td>
<td>-.021</td>
</tr>
<tr>
<td>54 Number of national scholastic honor societies in college.</td>
<td>.717</td>
<td>-.009</td>
<td>.011</td>
</tr>
<tr>
<td>24 Highest education level achieved.</td>
<td>.716</td>
<td>-.047</td>
<td>.040</td>
</tr>
<tr>
<td>42 College science grade received most often.</td>
<td>.694</td>
<td>.001</td>
<td>.039</td>
</tr>
<tr>
<td>41 College math grade received most often.</td>
<td>.691</td>
<td>-.016</td>
<td>.063</td>
</tr>
<tr>
<td>47 Percent of college expenses covered by scholastic scholarships</td>
<td>.497</td>
<td>-.026</td>
<td>.115</td>
</tr>
<tr>
<td>37 Number of college courses I failed...</td>
<td>.346</td>
<td>-.005</td>
<td>-.041</td>
</tr>
<tr>
<td>Item</td>
<td>G1</td>
<td>G2</td>
<td>G3</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my logical reasoning skills as...</td>
<td>-0.015</td>
<td>0.718</td>
<td>-0.028</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my analytical skills as...</td>
<td>0.043</td>
<td>0.673</td>
<td>0.026</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my ability to recall facts and details as...</td>
<td>-0.026</td>
<td>0.648</td>
<td>-0.007</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my ability to think on my feet as...</td>
<td>-0.072</td>
<td>0.637</td>
<td>-0.052</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my ability to do several jobs at once as...</td>
<td>-0.050</td>
<td>0.638</td>
<td>-0.051</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my problem solving skills as...</td>
<td>-0.048</td>
<td>0.624</td>
<td>-0.026</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my reading comprehension as...</td>
<td>0.043</td>
<td>0.617</td>
<td>-0.011</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my planning and organizing skills as...</td>
<td>0.012</td>
<td>0.610</td>
<td>-0.048</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my vocabulary as...</td>
<td>0.061</td>
<td>0.593</td>
<td>0.032</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my speed of reading as...</td>
<td>0.028</td>
<td>0.665</td>
<td>-0.011</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my writing skill as...</td>
<td>0.092</td>
<td>0.649</td>
<td>0.040</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my basic math skills as...</td>
<td>0.036</td>
<td>0.499</td>
<td>0.188</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would rate my ability to master assignments as...</td>
<td>-0.031</td>
<td>0.489</td>
<td>0.035</td>
</tr>
<tr>
<td>High school grade received most often.</td>
<td>-0.041</td>
<td>-0.047</td>
<td>0.854</td>
</tr>
<tr>
<td>Class standing in high school...</td>
<td>0.027</td>
<td>0.045</td>
<td>0.778</td>
</tr>
<tr>
<td>Number of times you made honor roll in high school</td>
<td>0.045</td>
<td>0.047</td>
<td>0.834</td>
</tr>
<tr>
<td>High school math grade most often received</td>
<td>-0.039</td>
<td>-0.032</td>
<td>0.836</td>
</tr>
<tr>
<td>High school science grade most often received</td>
<td>0.017</td>
<td>0.018</td>
<td>0.620</td>
</tr>
<tr>
<td>Relative to other high school students, my most demanding teacher would describe my academic work...</td>
<td>-0.122</td>
<td>0.135</td>
<td>0.616</td>
</tr>
<tr>
<td>High school English grade received most often.</td>
<td>0.025</td>
<td>0.077</td>
<td>0.571</td>
</tr>
<tr>
<td>Number of high school courses I failed.</td>
<td>0.026</td>
<td>0.091</td>
<td>0.459</td>
</tr>
<tr>
<td>High school teachers would most likely describe my academic potential as...</td>
<td>-0.026</td>
<td>0.283</td>
<td>0.379</td>
</tr>
<tr>
<td>When I graduated from high school I was (16,17,18,19, 20 years of age or older)...</td>
<td>0.026</td>
<td>0.031</td>
<td>0.043</td>
</tr>
</tbody>
</table>

α = 0.93

.78

.85
(Table 4.3 continued)

<table>
<thead>
<tr>
<th>Motivation Items</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = .38$</td>
</tr>
<tr>
<td>I was chosen to serve on special task forces or committees at work...</td>
<td>.606</td>
</tr>
<tr>
<td>I was chosen to serve as supervisor in my boss' absence...</td>
<td>.579</td>
</tr>
<tr>
<td>My rate of promotion was...</td>
<td>.535</td>
</tr>
<tr>
<td>I was selected to attend training...</td>
<td>.519</td>
</tr>
<tr>
<td>Number of formal awards I got for job performance...</td>
<td>.390</td>
</tr>
<tr>
<td>Time worked on my last full-time job...</td>
<td>.341</td>
</tr>
<tr>
<td>In the past 3 years, number of promotions I received on jobs was:</td>
<td>.422</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would describe my skill at meeting deadlines under pressure as:</td>
<td>.448</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would describe my skill at getting work done on time as:</td>
<td>.433</td>
</tr>
<tr>
<td>Prior to accepting my first job in my present job series, I worked extra hours on evenings or weekends...</td>
<td>.385</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would describe the amount of supervision that I need as...</td>
<td>.307</td>
</tr>
<tr>
<td>My previous supervisor (or teachers...) would describe the speed at which I work as...</td>
<td>.456</td>
</tr>
<tr>
<td>In my previous job, I was late (tardy for work):</td>
<td>.361</td>
</tr>
<tr>
<td>The amount of time I have been out of work between jobs usually has been:</td>
<td>.411</td>
</tr>
<tr>
<td>On my last job, my supervisor rated me as:</td>
<td>.490</td>
</tr>
<tr>
<td>Proportion of college expenses I earned.</td>
<td>.093</td>
</tr>
<tr>
<td>Average number of hours paid employment/week.</td>
<td>.088</td>
</tr>
<tr>
<td>The number of times I elected non-required college science courses was:</td>
<td>-.045</td>
</tr>
<tr>
<td>The number of times I elected non-required college math courses was:</td>
<td>-.021</td>
</tr>
<tr>
<td>The number of letters I received in college sports was:</td>
<td>-.073</td>
</tr>
<tr>
<td>The number of times I elected non-required college English courses was:</td>
<td>-.020</td>
</tr>
<tr>
<td>At the time I applied for this job, my undergraduate education consisted of having completed:</td>
<td>-.153</td>
</tr>
<tr>
<td>The number of high school clubs and organized activities in which I participated was:</td>
<td>.064</td>
</tr>
<tr>
<td>Prior to this job, amount of formal training (other than college) I participated related to my present job:</td>
<td>.109</td>
</tr>
<tr>
<td>The number of letters I received in high school sports was:</td>
<td>.061</td>
</tr>
<tr>
<td>My final year in high school, I was absent...</td>
<td>-.041</td>
</tr>
<tr>
<td>In the past 6 months, average number of hours/week I spent reading newspapers, books, outside of work:</td>
<td>.088</td>
</tr>
<tr>
<td>The number of letters I received in high school sports was:</td>
<td>.092</td>
</tr>
</tbody>
</table>
### Motivation Items (cont.)

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>Number of civic or social organizations (with regular meetings) I belonged to prior to accepting this job:</td>
<td>.056 .146</td>
</tr>
<tr>
<td>56</td>
<td>In the three years prior to accepting this job, number of different full- or part-time jobs I applied for was:</td>
<td>.149 -.137</td>
</tr>
<tr>
<td>4</td>
<td>During my last year in high school, my average number of hours of paid employment per week was:</td>
<td>.254 -.099</td>
</tr>
<tr>
<td>39</td>
<td>At the time I applied for this job, my graduate education consisted of having completed:</td>
<td>-.039 .076</td>
</tr>
<tr>
<td>83</td>
<td>My previous supervisor (or teachers...) would describe my attendance record as...</td>
<td>.283 .052</td>
</tr>
<tr>
<td>134</td>
<td>During my teens, I usually spent most of my summers (taking life easy...working full time):</td>
<td>.246 .044</td>
</tr>
<tr>
<td>91</td>
<td>My previous supervisor (or teachers...) would describe me as taking on more than I can handle:</td>
<td>-.085 -.040</td>
</tr>
<tr>
<td>74</td>
<td>In the year before accepting this job, the number of times I had been late for work (or class) was:</td>
<td>.184 -.035</td>
</tr>
<tr>
<td>73</td>
<td>The age at which I first started to earn money (other than an allowance) was:</td>
<td>.192 .029</td>
</tr>
<tr>
<td>71</td>
<td>The amount of time I have been out of work between jobs usually has been...</td>
<td>.267 .024</td>
</tr>
<tr>
<td>61</td>
<td>Prior to accepting this job, the number of different federal agencies I worked for was:</td>
<td>-.060 -.011</td>
</tr>
<tr>
<td>75</td>
<td>In the three years prior to accepting this job, the number of jobs I had been fired from was...</td>
<td>.120 -.006</td>
</tr>
<tr>
<td>89</td>
<td>My previous supervisor (or teachers...) would describe amount of time needed to complete assignments as</td>
<td>.280 .002</td>
</tr>
<tr>
<td>65</td>
<td>The number of months I was unemployed during the 3 years immediately prior to this job...</td>
<td>.223 .000</td>
</tr>
</tbody>
</table>

### Self Perception Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>114</td>
<td>On a list of 100 people in the kind of job I can do best, my peers would place me in the (top 10, 25, 50, 75, 90%).</td>
<td>.76</td>
</tr>
<tr>
<td>117</td>
<td>My peers would probably say that the highest level I could reach if I chose a career in major corporation would be...</td>
<td>.82</td>
</tr>
</tbody>
</table>

---

1 Items used with permission of the Federal Aviation Administration (see Appendix)

* Shaded factor loadings represent item parcels used in confirmatory factor analysis.

** Cronbach's alphas of item parcels

*** Cronbach's alphas of Q-sort item grouping
account the error of approximation in the population and asks the question, "How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?" (Browne & Cudeck, 1993, pp. 137-138).

Confirmatory factor analysis results suggested a priori Q-sort item groupings consistent with the ecology model (Model 1) did not adequately fit the data (Model 1: $\chi^2 [34, N = 6036] = 8145.78$, $p < .001$; GFI = .81, CFI = .56, NFI = .56, PNFI = .35, and RMSEA = .19). Fit indices were not acceptable using commonly used heuristics in the literature (e.g., Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989).

An attempt was made to improve fit using the Q-sort common factor analysis results and initial CFA modification indices. A sequence of rational exploratory analyses and confirmatory factor analyses, or an Iterative Rational Empirical (IRE) approach, was used to examine other latent structures. IRE describes post hoc interpretation of exploratory common factor analyses within the five ecology model-based construct domains to alter measurement models examined in subsequent CFA.

Common factor analysis results suggested different "time windows" of developmental opportunity may exist within each ecology model construct domain (Rovee-Collier, 1995). For example, factor loadings tapping evidence of cognitive ability seem to reflect high school, college, and on-the-job developmental time periods. Further, factors appear to reflect different
perspectives or views through these developmental windows (i.e., self, peer, superior, teacher, friend, and co-worker views).

Rovee-Collier (1995) proposed time-windows as a key concept in cognitive development. A time-window is a critical period where information about a current event is integrated with previously acquired information. However, if the same information is encountered outside of the time-window, it will not be integrated. Time-windows are not restricted to a particular age or stage of development. Nonetheless, they are open for a limited duration before closing. Discrete events occurring outside of a time-window are treated as unique and thus are not assimilated into the reservoir of collective memory. Rovee-Collier (1995) speculated time-windows may be the cornerstone of individual differences in cognitive domains involving integration of successive experiences. She asserted that as personal experiences of same-age individuals differ from moment to moment, so will their developmental time-windows, what they remember from those time-windows, and whether the new information will be integrated with existing information in the future. Alternatively, developmental negative life events that occur when time-windows are closed may be more likely to produce intensified distress upon re-exposure.

Based on prior findings and theory from Rovee-Collier (1995), a rational approach was used to group these parcels according to the time windows captured by biodata items within that parcel. The parcels seemed to fit into
developmental windows occurring on the job, during high school, and in college. Parcels representing developmental job experiences included: Interpersonal Parcel 1 (I1; My previous supervisor {or teachers if not previously employed} would rate my oral communication skills as...), Personality Parcel 1 (P1; My previous supervisor {or teachers if not previously employed} would rate my dependability as...), Cognitive Ability Parcel (g2; My previous supervisor {or teachers if not previously employed} would rate my logical reasoning skills as...), Motivation Parcel 1 (M1; I was chosen to serve on special task forces or committees at work...), and Self Perception Parcels 1 and 2 (P1; On a list of 100 people in the kind of job I can do best, my peers would place me in the top 10, 25, 50, 75, 90%; P2; My peers would probably say that the highest level I could reach if I chose a career in major corporation would be...).

Another group of parcels seemed to tap college experiences: Cognitive Ability Parcel 1 (g1; Number of times you made the Dean’s List in college...), and Motivation Parcel 2 (M2; The number of times I elected non-required college math courses was...). Finally, a group of parcels seemed to tap high school experiences: Interpersonal Parcel 2 (I2; Relative to other high school students, my classmates would most likely describe my interpersonal skills as...), Cognitive Ability Parcel 3 (g3; Number of high school courses I failed...). Personality Parcel 2 (P2; My peers would describe my aggressiveness as...)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
addressed peers' views of applicant personality, not directly addressing a particular situation or point in time.

The second CFA model examined a three-factor time windows framework (Job, College, High School). The indicators of each factor for this model were as follows: Job: l1, P1, g2, M1, S1, & S2; College: g1, M2; High School: l2, g3, P2. This model was derived from the rational interpretation of exploratory analyses suggesting data may fit the time-windows based model. Hence, parcels were rationally loaded onto latent time windows constructs.

This model fit the data better than ecology model derived Model 1 (Model 2: $\chi^2 \ [41, N = 6036] = 2148.22, p < .001; GFI = .94, CFI = .89, NFI = .88, PNFI = .58, and RMSEA = .092$). Lambda-x modification indices suggested Interpersonal Parcel 2 (l2) be loaded onto High school instead of College. Model 3 found this change to yield a small increase in quality of fit (Model 3: $\chi^2 \ [41, N = 6036] = 1815.37, p < .001; GFI = .95, CFI = .90, NFI = .90, PNFI = .55, and RMSEA = .085$). A low P2 path coefficient (.05) and low lambda-x modification index indicated moving this path to another factor would not improve model fit.

Recall items in Personality parcel 2 were not specific to a particular period of time window. Hence, the fact that Personality Parcel 2 did not load onto any latent time windows constructs is consistent with that parcel’s broad item content. Model 4 reflects the deletion of the P2 indicator from the model. Model 4 also fit relatively well (Model 4: $\chi^2 \ [32, N = 6036] = 1740.56, p < .001; GFI = .94, CFI = .91, NFI = .91, PNFI = .55, and RMSEA = .094$).
High error terms on the Self-Perception parcels relative to other indicators was a concern but was explained by the fact that each parcel contained only one item. Self-perception parcels were subsequently removed in Model 5, generating acceptable fit indices (Model 5: $\chi^2 [17, N = 6036] = 1347.62$, $p < .001$; GFI = .95, CFI = .92, NFI = .92, PNFI = .45, and RMSEA = .11). Finally, College and High School factors were highly correlated (.73), and a two-factor model (Job and School) was submitted. Model 6 fit did not improve over other models examined (Model 6: $\chi^2 [19, N = 6036] = 2104.32$, $p < .001$; GFI = .92, CFI = .88, NFI = .87, PNFI = .49, and RMSEA = .13). Table 4.4 contains a summary of fit indices and rationales for all models.

Results bearing on Research Question 1 did not provide strong initial support for the ecology model. Model fit was achieved by grouping ecology model-based indicators according to the time windows rationale. A series of slight modifications to the time windows model did not greatly affect the already high levels of goodness of fit. No one time windows model seemed to greatly surpass another, as all had consistently high fit indices and low RMSEA. Models 3 and 5 seem to edge out other models in terms of data fit. Model 3 may be the better of the two because it had the lowest RMSEA.

Research Question 2 was addressed by assessing the predictive validity of Q-sorted item groupings using a response option-based empirical key. Subject response options (0 = non selected, 1 = selected) in the cross-validation sample were multiplied by each option's point biserial correlation.
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Change</th>
<th>Reason for change</th>
<th>$\chi^2$</th>
<th>df</th>
<th>GFI</th>
<th>CFI</th>
<th>NFI</th>
<th>PGFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5-factor Ecology Model</td>
<td></td>
<td></td>
<td>8145.78</td>
<td>34</td>
<td>.81</td>
<td>.19</td>
<td>.56</td>
<td>.56</td>
<td>.35</td>
</tr>
<tr>
<td>2</td>
<td>Initial 3 Factor Time-Windows Ecology Model</td>
<td>Ecology model fit to time-windows</td>
<td>Low GFIs for the Ecology Model</td>
<td>2148.22</td>
<td>41</td>
<td>.94</td>
<td>.89</td>
<td>.88</td>
<td>.58</td>
<td>.092</td>
</tr>
<tr>
<td>3</td>
<td>Modified 3-factor Time-Windows Ecology Model</td>
<td>I2 moved from HS to College factor</td>
<td>High lambda-x modification index for I2 on College factor</td>
<td>1815.37</td>
<td>41</td>
<td>.95</td>
<td>.90</td>
<td>.90</td>
<td>.59</td>
<td>.085</td>
</tr>
<tr>
<td>4</td>
<td>Modified 3-factor Time-Windows Ecology Model</td>
<td>P2 dropped from the model</td>
<td>Low P2 path loading on HS &amp; low modification indices for P2 on other factors</td>
<td>1740.56</td>
<td>32</td>
<td>.94</td>
<td>.91</td>
<td>.91</td>
<td>.55</td>
<td>.094</td>
</tr>
<tr>
<td>5</td>
<td>Modified 3-factor Time-Windows Ecology Model</td>
<td>S1 &amp; S2 dropped</td>
<td>High measurement error in S1 &amp; S2</td>
<td>1347.62</td>
<td>17</td>
<td>.95</td>
<td>.92</td>
<td>.92</td>
<td>.45</td>
<td>.11</td>
</tr>
<tr>
<td>6</td>
<td>2-factor Time-Windows Ecology model</td>
<td>College and High school factor merged into one factor</td>
<td>College and High school factors highly correlated</td>
<td>2104.32</td>
<td>19</td>
<td>.92</td>
<td>.88</td>
<td>.87</td>
<td>.49</td>
<td>.13</td>
</tr>
</tbody>
</table>

*Iterative-Rational-Empirical (IRE) describes sequences of interpreting 1) exploratory common factor analyses loadings within the five ecology model-based construct domains designed to generate more homogeneous item parcels and 2) CFA modification indices from the original 5 factor model in order to specify alternate measurement models.

**GFI = goodness of fit index, AGFI = adjusted goodness of fit index, CFI = comparative fit index, NFI = normed fit index, PNFI = parsimonious normed fit index, RMSEA = root mean squared error of approximation.
with the criterion (the FAA Academy composite score) in the key development sample and summed to yield scale scores for biodata item groupings. Criterion validity results for these scales are found in column 1 of Table 4.2. Cross-validities for the ecology construct categories of interpersonal, personality, cognitive ability, motivation, and self-perception were .138, .124, .296, .209, and .064, respectively, and generally moderate to low.

The moderate to low inter-correlations among ecology model scales constitute evidence of discriminant validity. Highest inter-correlations were between Personality and Motivation (.436), Cognitive ability and Self-perception (.301), Self-Perception and Motivation (.273), Cognitive ability and Motivation (.269), and Cognitive ability and Personality (.268). The Interpersonal scale did not correlate meaningfully with any other scale. The correlation between Motivation and Personality could be viewed as evidence of convergent validity, as theory suggests one aspect of personality is "conscientiousness" which is conceptually similar to the "motivation" construct (Barrick & Mount, 1991). Motivation and Cognitive ability were also expected to be moderately correlated because both typically need be present for performance to occur (Campbell, Dunnette, Lawler, & Weick, 1970) and this was a relatively range restricted, high performing sample. These inter-correlations suggested support for convergent and divergent validities of these scales consistent with theory-based expectations.
Research Question 3 examined the relative criterion validity of biodata items that could and could not be assigned to theory-based categories. The criterion validity of eleven items not sorted into theory-based construct domains was compared to five randomly selected groups of eleven theory-based biodata items. The cross-validity obtained for non-theory based items was .128 while cross-validities obtained for the random groups of theory-based items were .157, .243, .194, .166, and .169 with an average cross-validity of .186. Using the Hotelling-Williams test (Bobko, 1995; Williams, 1959), the hypothesis \( H_0: \rho_{yx} = \rho_{yz} \) was tested to determine whether there was a significant difference between the two dependent correlations. The correlations were not independent of each other because they were computed on the same sample and had a common dependent variable (Bobko, 1995).

The Hotelling-Williams test yielded \( t_{672} = 1.4617 \), which was non-significant in a 1-tailed test (critical value = 1.645). Hence, difference between the cross-validities of the theory based items versus non-theory based items was not statistically significant. No evidence was found to suggest theory based items provide higher criterion-related validities.

In sum, there was mixed support for construct validity of this biodata instrument based on the ecology model framework. The biodata items were reliably Q-sorted into ecology model constructs as evidenced by relatively high inter-rater agreement. Correlations among Q-sorted ecology scales demonstrated discriminant validity among the scales. Parcels used for
confirmatory factor analyses of the Q-sort groupings also had relatively high alphas (ranging from .58 - .93) given the parcels were based on the item Q-sort and not based on items developed with ecology model constructs in mind.

Biodata Criterion Validity

The second set of analyses addressed biodata criterion-related validity. The following specific research questions were posed:

Research Question #4: Is the biodata measure a valid predictor of performance?

Research Question #5: Is the general cognitive ability measure a valid predictor of performance?

Research Question #6: What are the relative contributions of biodata and general cognitive ability measures to performance prediction?

Research Question #7: Are 1) “g-loaded” and 2) “non-g-loaded” biodata items differentially related to the general cognitive ability measure?

Research Question #8: Do “g-loaded” biodata items, “non-g-loaded” biodata items, and the cognitive ability test exhibit incremental validity relative to one another?

It was necessary to first determine that both instruments were indeed valid predictors. Biodata and cognitive ability measures were both correlated with the criterion to determine each instrument’s criterion-related validity and
address Research Questions 4 and 5. Table 4.5 presents results of cross-validities obtained for the biodata instrument, two biodata sub-scales (g items and non-g items), and the general cognitive ability test with Screen performance.

Table 4.5
Simple Correlations (cross-validities)*

<table>
<thead>
<tr>
<th></th>
<th>1. Criterion</th>
<th>2. Biodata</th>
<th>3. Biodata (Bg)</th>
<th>4. Biodata (B_non-g)</th>
<th>5. g measure</th>
<th>6. HS scale</th>
<th>7. College scale</th>
<th>8. Job scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>.365 (.44)**</td>
<td>.295 (.36)**</td>
<td>.269 (.29)**</td>
<td>.155 (.42)***</td>
<td>.202 (.32)**</td>
<td>.194 (.19)**</td>
<td>.188 (.17)**</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>.132</td>
<td>.694</td>
<td>.582</td>
<td>.483</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.139</td>
<td>.806</td>
<td>.593</td>
<td>.296</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.037</td>
<td>.045</td>
<td>.200</td>
<td>.613</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Note – all are significant at p < .001, N >= 748
** Correlation corrected for indirect range restriction on the cognitive ability measure
*** Correlation corrected for direct range restriction on the cognitive ability measure

The biodata instrument correlated .365 (.44, corrected for indirect range restriction) and the cognitive ability test correlated .155 (.42, corrected for direct range restriction) with Screen performance. Direct range restriction occurred due to the fact that the general cognitive ability test was used to select applicants, hence any correlation between g and the criterion was attenuated by loss of the low end of the g distribution. Correction for indirect range restriction on the biodata criterion validity adjusts for the fact that subjects were selected (and hence range restricted) on general cognitive ability (Bobko, 1995). These correlations suggest the two instruments were valid predictors of performance and justified further analyses to determine the instruments incremental validities.
Relative contributions of biodata and cognitive ability to performance prediction was assessed using hierarchical multiple regression using a matrix of simple correlations corrected for range restriction on the general cognitive ability measure as input. Results of these analyses are summarized in Table 4.6. Biodata is the more powerful predictor based on uncorrected simple correlations (.365 v. .155), simple correlations corrected for direct and indirect range restriction (.440 v. .420), and incremental predictive power (ΔR).

Biodata yielded ΔR of .113 when added to a regression equation with g, while g yielded ΔR of .071 when added to a regression equation containing biodata.

Table 4.6
Hierarchical Multiple Regression Analysis Results

<table>
<thead>
<tr>
<th>Variable(s) Entered</th>
<th>N</th>
<th>R</th>
<th>ΔR</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>1893</td>
<td>.395</td>
<td></td>
</tr>
<tr>
<td>Biodata</td>
<td>686</td>
<td>.437</td>
<td></td>
</tr>
<tr>
<td>Biodata_g</td>
<td>768</td>
<td>.376</td>
<td></td>
</tr>
<tr>
<td>Biodata_non-g</td>
<td>718</td>
<td>.277</td>
<td></td>
</tr>
<tr>
<td>g + Biodata</td>
<td>686</td>
<td>.508</td>
<td>.113</td>
</tr>
<tr>
<td>g + Biodata_g</td>
<td>768</td>
<td>.501</td>
<td>.106</td>
</tr>
<tr>
<td>g + Biodata_non-g</td>
<td>718</td>
<td>.468</td>
<td>.073</td>
</tr>
</tbody>
</table>

Research Question 7 examined whether "g-loaded" and "non-g-loaded" biodata items were differentially related to the general cognitive ability measure. Non-g-loaded biodata items were made up of all items that were not Q-sorted as overlapping dominantly with the general cognitive ability construct domain (i.e., interpersonal, personality, motivation, self-perception, and non-theory-based items). Some evidence of convergent/discriminant validity was
demonstrated as the g loaded items and the non-g items correlated with the general cognitive ability test \( r = .139 \) (\( r_c = .345 \)) and \( r = .037 \) (\( r_c = .100 \)), respectively. The g-loaded items were more highly correlated than the non-g biodata items with the general cognitive ability test, as expected. Overall correlations with the general cognitive ability measure were relatively small, suggesting g-loaded biodata items were capturing g as well as other constructs.

Additional analyses were conducted to examine simple criterion validities of the high school, college, and job scales with the general cognitive ability measure. These scales emerged from exploratory analyses that examined the fit of biodata theory in the current data set. Results are presented in Table 4.5 above. The high school, college and job scales were correlated with the criterion .202 (\( r_c = .32 \)), .206 (\( r_c = .19 \)), and .170 (\( r_c = .17 \)), respectively. Corrected correlations were adjusted for indirect range restriction on the general cognitive ability measure.

Research Question 8 additionally asked whether “g-loaded” biodata items, “non-g-loaded” biodata items, and the general cognitive ability measure exhibit incremental validity relative to one another. Hierarchical multiple regression using a matrix of corrected simple correlations as input was used to examine incremental validities. The g-loaded biodata items correlated .376 (corrected for indirect range restriction on g) with the criterion and incrementally yielded \( \Delta R \) of .106 when added to the general cognitive ability
test. The non-g-loaded biodata items correlated .270 (corrected for indirect range restriction on g) with the criterion and yielded $\Delta R$ of .073 when added to the general cognitive ability test. The general cognitive ability test added to the g-loaded biodata items yielded $\Delta R$ of .125. The general cognitive ability test was added to the non-g loaded biodata items yielded $\Delta R$ of .191. These results indicate the entire biodata scale adds more to prediction than the general cognitive ability test. When the scale was divided into g and non-g components, the general cognitive ability test outperformed the biodata subscales in terms of incremental criterion validity. The general cognitive ability test yielded higher incremental validity with the biodata non-g scale than with the g scale.

Additional analyses examined incremental criterion validity of the set of biodata scales (high school, college, and job) with the general cognitive ability test. The $\Delta R$s were obtained from hierarchical multiple regression using simple correlations corrected for range restriction on the general cognitive ability measure. These scales emerged from the exploratory analyses addressing biodata theory issues. Results are presented in Table 4.7.

Biodata high school items yielded $\Delta R$ of .083 when added to the general cognitive ability measure. When general cognitive ability measure was added to high school biodata items, $\Delta R$ was .163. College items added incrementally to prediction with the general cognitive ability test with $\Delta R = .062$. Conversely, when the general cognitive ability test was added to an equation with the
Table 4.7
Hierarchical Multiple Regression Analysis Results: Exploratory Biodata Scales

<table>
<thead>
<tr>
<th>Variable(s) Entered</th>
<th>N</th>
<th>R</th>
<th>ΔR</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>1893</td>
<td>.395</td>
<td></td>
</tr>
<tr>
<td>HS scale</td>
<td>825</td>
<td>.315</td>
<td></td>
</tr>
<tr>
<td>College scale</td>
<td>799</td>
<td>.189</td>
<td></td>
</tr>
<tr>
<td>Job scale</td>
<td>823</td>
<td>.173</td>
<td></td>
</tr>
<tr>
<td>g + HS scale</td>
<td>825</td>
<td>.478</td>
<td>.478 - .395 = .083</td>
</tr>
<tr>
<td>g + College scale</td>
<td>799</td>
<td>.457</td>
<td>.457 - .395 = .062</td>
</tr>
<tr>
<td>g + Job scale</td>
<td>823</td>
<td>.451</td>
<td>.451 - .395 = .056</td>
</tr>
</tbody>
</table>

college biodata items, ΔR = .268. Finally, the job biodata items yielded an incremental validity of .056 with the cognitive ability measure; the general cognitive ability test yielded an incremental validity of .279 when added to the job biodata items.

In sum, the entire biodata scale outperformed the general cognitive ability test in terms of uncorrected and corrected criterion validity. Using corrected correlations to assess incremental criterion validity, the entire biodata scale outperformed the general cognitive ability test, however, the general cognitive ability test outperformed the biodata sub-scales (g item and non-g item scales). High school, college, and job biodata item scale correlations with performance were interesting as the high school items (rc = .32) outperformed the college and job items and the college items (rc = .19) outperformed the job items (rc = .17) in terms of criterion-related validity.

Surprisingly, the more temporally removed from performance the biodata scale was, the more it predicted future job performance. The general cognitive
ability test outperformed the three biodata scales in terms of incremental validity adding the most to prediction when combined with the job biodata scale ($\Delta R = .279$), adding least when combined with the high school biodata scale ($\Delta R = .163$).

**Biodata Adverse Impact**

The final set of analyses addressed adverse impact of biodata and general cognitive ability measures:

**Research Question #9:** When adverse impact response options are removed from the empirical key, is there a significant change in biodata criterion-related validity and adverse impact?

**Research Question #10:** Do general cognitive ability and biodata measures differ significantly in terms of adverse impact?

The first analysis examined whether response options were chosen with different frequency across subgroups. Each response option was examined against the Equal Employment Opportunity Commission’s four-fifths rule to determine whether African Americans answered response options with differential frequency (i.e., at a rate less than 80% or more than 120% of the majority group).

Of the 710 biodata response options, 129 were selected with differentially lower frequency by blacks and 144 by whites. Criterion-related validity of the biodata inventory was analyzed with these adverse impact
response options excluded to address Research Question 10. This analysis was consistent with EEOC Uniform Guidelines (1978) requirements that selection systems not be analyzed at the component level but rather at the level of the overall effect. Uncorrected cross-validity obtained with all adverse impact response options excluded was .344. This compares to an uncorrected cross-validity of .365 of the entire biodata inventory with the criterion.

A Hotelling-Williams Test was performed to determine whether the correlation between the biodata instrument and the criterion was significantly different from the correlation obtained when adverse impact response options were removed from the instrument. The difference was not significant ($t_{818} = .1605$, critical value = 1.645 for a one-tailed test). Interestingly, the number of adverse impact response options for blacks and whites was very similar, suggesting their net combined effect on an overall biodata score would be negligible. Indeed, the small validity decrement suggested this to be the case. This finding is consistent with previous findings indicating biodata tends to have very low adverse impact on minority groups (Reilly & Chao, 1982; Reilly & Warech, 1990).

The biodata instrument (including and excluding adverse impact response options) and general cognitive ability measure were then tested to see if either violated the Cleary (1968) model of test bias (also referred to as the regression model by the EEOC Uniform Guidelines, 1978). The Cleary model states:
"A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup" (Cleary, 1968, p. 115).

The fairness of the biodata instrument (with and without adverse impact response options) and cognitive ability test were tested by running moderated multiple regression for each instrument as follows:

\[ Y_{predicted} = constant + selection device + race + selection device \times race \]

Moderated regression results are presented in Table 4.8. None of the interaction terms were statistically significant from zero, hence all of the predictors exhibited test fairness as per the Cleary model. Maxwell and Arvey (1993) demonstrated that “within the universe of fair tests (as defined by T. A. Cleary, 1968), the most valid selection method will necessarily produce less adverse impact” (p. 433). Hence, the biodata instrument is preferred on both of the Uniform Guidelines' (1978) double hurdles of adverse impact and test fairness. The biodata scales (including all response options and including only non-adverse impact response options) were expected to demonstrate less adverse impact because they yielded higher criterion validities \((r_c = .427 \text{ and } .410, \text{ respectively})\) than the criterion validity obtained from the general cognitive ability test \((r_c = .395)\). It is noted that Maxwell and Arvey’s proof applies to independent measures so it does not apply when comparing non-
independent biodata scales (e.g., the full biodata and non-adverse impact
biodata scales).

Table 4.8
Analysis of Test Fairness: Cleary Model of Test Bias

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variables Entered</th>
<th>B</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive ability test</td>
<td>Cognitive ability test</td>
<td>.607</td>
<td>4.746</td>
<td>.000</td>
</tr>
<tr>
<td>N = 8838</td>
<td>Race</td>
<td>8.641</td>
<td>.079</td>
<td>.427</td>
</tr>
<tr>
<td>R² = .228</td>
<td>Cognitive ability*Race</td>
<td>-.018</td>
<td>-1.520</td>
<td>.131</td>
</tr>
<tr>
<td>Biodata</td>
<td>Biodata</td>
<td>2.681</td>
<td>2.215</td>
<td>.027</td>
</tr>
<tr>
<td>N = 3414</td>
<td>Race</td>
<td>-107.07</td>
<td>-.910</td>
<td>.363</td>
</tr>
<tr>
<td>R² = .368</td>
<td>Biodata*Race</td>
<td>1.002</td>
<td>0.854</td>
<td>.393</td>
</tr>
<tr>
<td>Biodata (non-AI response options only)</td>
<td>Biodata_{Non-AI}</td>
<td>5.138</td>
<td>2.958</td>
<td>.003</td>
</tr>
<tr>
<td>N = 3423</td>
<td>Race</td>
<td>-16.433</td>
<td>-.098</td>
<td>.922</td>
</tr>
<tr>
<td>R² = .357</td>
<td>Biodata_{Non-AI}*Race</td>
<td>.088</td>
<td>.052</td>
<td>.958</td>
</tr>
</tbody>
</table>

A final set of analyses addressed Research Question 10 and the degree
of adverse impact exhibited by biodata and the general cognitive ability test.
Cut scores at the 20th, 40th, 60th, and 80th percentiles were used to assess
adverse impact for the cognitive ability and biodata instruments. These
arbitrary cut points were meant to be illustrative and are only used because the
sample had been pre-screened, i.e., only those applicants with cognitive ability
test scores greater than 90 were actually hired. As criterion data was not
available for the entire applicant pool, adverse impact at cut points lower than
90 could not be examined. Hence, the adverse impact (or lack thereof) found
in this sample cannot be viewed as representative of adverse impact that might occur at cut scores falling outside the current sample range.

The ratio of percent blacks versus whites that would have been hired using the cognitive ability test, biodata instrument, or biodata instrument with adverse impact response options removed was examined at each cut score and compared to the EEOC's four-fifths rule. Results are reported in Table 4.9.

Table 4.9
Adverse Impact Analyses

<table>
<thead>
<tr>
<th>Cut Score (percentiles)</th>
<th>Selection Device</th>
<th>Black Selection Ratio</th>
<th>White Selection Ratio</th>
<th>Adverse Impact Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>80th</td>
<td>Cognitive Ability</td>
<td>.16</td>
<td>.22</td>
<td>.70**</td>
</tr>
<tr>
<td></td>
<td>Biodata</td>
<td>.10</td>
<td>.22</td>
<td>.44**</td>
</tr>
<tr>
<td></td>
<td>Biodata&lt;sub&gt;Non-Al&lt;/sub&gt;</td>
<td>.18</td>
<td>.22</td>
<td>.80</td>
</tr>
<tr>
<td>60th</td>
<td>Cognitive Ability</td>
<td>.27</td>
<td>.36</td>
<td>.77**</td>
</tr>
<tr>
<td></td>
<td>Biodata</td>
<td>.30</td>
<td>.43</td>
<td>.70**</td>
</tr>
<tr>
<td></td>
<td>Biodata&lt;sub&gt;Non-Al&lt;/sub&gt;</td>
<td>.40</td>
<td>.43</td>
<td>.93</td>
</tr>
<tr>
<td>40th</td>
<td>Cognitive Ability</td>
<td>.35</td>
<td>.51</td>
<td>.69**</td>
</tr>
<tr>
<td></td>
<td>Biodata</td>
<td>.54</td>
<td>.63</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>Biodata&lt;sub&gt;Non-Al&lt;/sub&gt;</td>
<td>.66</td>
<td>.63</td>
<td>.96</td>
</tr>
<tr>
<td>20th</td>
<td>Cognitive Ability</td>
<td>.53</td>
<td>.74</td>
<td>.72**</td>
</tr>
<tr>
<td></td>
<td>Biodata</td>
<td>.82</td>
<td>.75</td>
<td>1.05****</td>
</tr>
<tr>
<td></td>
<td>Biodata&lt;sub&gt;Non-Al&lt;/sub&gt;</td>
<td>.85</td>
<td>.81</td>
<td>1.05****</td>
</tr>
</tbody>
</table>

* Adverse Impact ratio = minority selection ratio/majority selection ratio
** Adverse impact occurred for the given selection device and cut score
*** Biodata<sub>Non-Al</sub> = biodata instrument with all response options with differential response frequencies by race
**** Black selection ratio was in the majority.

At the 80<sup>th</sup> and 60<sup>th</sup> percentiles, the cognitive ability test and biodata instrument exhibited adverse impact against blacks while the biodata inventory with adverse impact items removed did not. At the 40<sup>th</sup> and 20<sup>th</sup> percentile cut scores, the general cognitive ability test continued to exhibit adverse impact.
against whites while both versions of the biodata instrument did not. In fact, at
the 20th percentile, blacks were selected at a rate higher than that of whites for
both versions of the biodata instrument though not at a rate that constituted
adverse impact against whites. This illustration is consistent with previous
findings that biodata tends to display less adverse impact than tests of general
cognitive ability (Reilly & Chao, 1982; Reilly & Warech, 1990).

Again, Maxwell and Arvey’s proof applies to independent tests. Hence,
as the biodata scale scores with and without adverse impact response options
are clearly not independent, these findings are also consistent with Maxwell
and Arvey’s proof. The independent tests (entire biodata scale v. general
cognitive ability test and the non-adverse impact biodata scale v. general
cognitive ability test) demonstrated that the biodata scales, which had higher
criterion validity than the general cognitive ability test, indeed had the least
adverse impact. Importantly, the general cognitive ability test did not pass the
four-fifths test at any cut score. These findings also suggest discarding those
response options which violate the 4/5th rule enabled biodata to pass the
adverse impact analysis at the all percentile cut scores.

In sum, this study found a large percentage of the biodata response
options exhibited adverse impact. When these response options were
removed, the standardized mean subgroup difference decreased by 66% yet
criterion-related validity decreased by only 5.7%. Both biodata scales and the
general cognitive ability test passed the Cleary model of test fairness and
exhibited similar corrected criterion validities. When the three predictors were examined for adverse impact in this data with percentile selection cut scores, the biodata scale without adverse impact items outperformed the general cognitive ability test and the biodata scale including all response options.
CHAPTER 5

DISCUSSION AND CONCLUSION

The purpose of this study was threefold. First, construct validity of biodata was examined to determine if biodata theory was useful in explaining biodata's often cited, but not well understood, strong criterion validity. Second, biodata was examined in terms of incremental criterion-related validity relative to a general cognitive ability test. The biodata instrument was also investigated in terms of criterion and incremental validity of two biodata predictor scales used in combination with a general cognitive ability test. Predictor scales consisted of “g-loaded” and “non-g-loaded” response options, respectively. Finally, biodata adverse impact was assessed in two ways. First, individual biodata response options were examined for possible adverse impact. Second, adverse impact of separate biodata scales including and excluding adverse impact response options were compared to a test of general cognitive ability. Research findings are discussed and implications for theory, future research, and practice are offered.

Biodata Construct Validity

The first issue addressed was biodata construct validity. The ecology model (Mumford & Stokes, 1992; Mumford, Stokes, & Owens, 1990) served as the theoretical basis for this assessment. Construct validity of biodata was investigated using expert judgement and statistical assessment of content validity, convergent and discriminant validities, internal consistency reliability,
confirmatory factor analysis, and criterion-related validity. Research questions 1 through 3 addressed issues relating to construct validity.

**Research Question 1.** This research question asked whether biodata items displayed psychometric characteristics (e.g., content validity, internal consistency reliability, and item factor analytic loadings) consistent with theory-based construct domains. Results indicated partial support for the construct validity of this instrument using construct domains drawn from the ecology model. Q-sort results of biodata items onto ecology model construct domains yielded relatively high initial inter-rater agreement (82%). This level of agreement on ecology model construct domains was respectable given items were not developed with these constructs in mind. Group consensus discussion among raters was used to reach agreement on all 142 items. A majority of items (92%) were sorted into an ecology model domain. High inter-rater agreement coupled with majority of items being sorted into ecology model framework provided initial evidence for content validity of the biodata instrument.

Initial internal consistency reliabilities of the items Q-sorted into ecology model construct domains were at low to moderate levels (ranging between $\alpha = .27 - .74$, average $\alpha = .46$). Exploratory factor analysis within ecology model construct domain yielded clean loadings for subsequent item parcel construction. Average factor loading on dominant factors was .58, while the average factor loading on non-dominant factors was .05. Even though factor
loadings were clean, confirmatory factor analysis of the ecology model yielded sub-optimal fit. Lack of fit could have been due to a number of causes. For example, the biodata instrument may not have adequately sampled the ecology model construct domains. This remains a viable explanation because the ecology model was applied to this instrument post hoc. Additionally, two of the five construct domains, personality and self-perception, had few items sorted into them (10 and 2 items, respectively) compared to the number of items sorted into the other groupings (interpersonal, cognitive ability, and motivation each having 21, 37, and 42 items, respectively). A better test of the ecology model would have equal and larger numbers of items per construct domain. It would have been unreasonable, however, to mandate Q-sorters create construct groupings with equal numbers of items as the goal of the Q-sort was to attempt to accurately group items by construct.

An equally viable (and probably better) explanation for ecology model sub-optimal fit is that the model may need further development to explain biodata predictive ability. This explanation is suggested by exploratory factor analyses within each of the five Q-sort factors that yielded interpretable sub-categories, or parcels. Additionally, most parcels had higher coefficient alphas ($\alpha = .23 - .93$, average $\alpha = .72$) than alphas generated from the initial five ecology model-based factors (average $\alpha = .46$), with lower alphas typically occurring on parcels with few items (e.g., a personality parcel had only 3 items, $\alpha = .23$).
The parcel factors may capture a new set of interpretable life history events focusing on developmental life periods or windows – high school, college, and job experiences. For example, within the general cognitive ability Q-sort factor, exploratory factor analysis suggested individuals answered these biodata items differentially based on the time-window each item addressed.

The fact that items were answered differentially by respondents across time windows suggested they evolved and changed over time, as posited by the ecology model. The ecology model’s temporal aspect suggests individuals learn and modify their behavior due to previous choices and situations encountered. These differences in within-construct, across-time measures speak to the importance of writing items that sample multiple developmental life periods. This instrument included only high school, college, and early career events. This limitation of items to only early life events (rather than later life events, e.g., mid-life/career events) could be deemed appropriate as applicants for the air traffic controller specialist position were age 30 or younger as per congressional mandate.

The psychological construct of time-windows proposed by Rovee-Collier (1995) captures critical periods where information about current events is easily integrated with information and knowledge acquired from previous events to generate learning and development. Time-windows are not necessarily restricted to a particular age or stage of development, though high school, college, and early career periods would seem likely candidates as key
generalized developmental time windows. Time windows could be used to augment the ecology model by providing a richer description of the process by which individuals develop and evolve over time.

Another public sector biodata study interestingly found similar factor analysis results. Gandy, Dye, and MacLane (1992) factor analyzed the Individual Achievement Record, an 84-item biodata form used for federal agency entry-level positions. Gandy et al. (1992) reported four underlying factors: work competency, high school achievement, college achievement, and leadership skills. These findings suggested there was something unique to be learned by tapping a construct such as general cognitive ability across many different life events.

Regardless, both explanations for the ecology model sub-optimal fit (applying theory post hoc to the biodata instrument or that the theory may need more development) is equally viable. Current study design and data cannot address which is more correct.

Research Question 2. The second research question examined whether relationships among scale scores derived for latent biodata constructs yielded convergent and discriminant validities consistent with theory-based construct domains. Results indicated most correlations among the five factors were at low to moderate levels. The low correlations suggested each Q-sorted group of items based on ecology model construct domains tapped unique constructs.
Three moderate correlations among Q-sorted factors were found between motivation and personality; motivation and general cognitive ability; general cognitive ability and personality. The correlation between motivation and personality could be evidence of convergent validity, as one aspect of personality is “conscientiousness” (Barrick & Mount, 1991) which may have conceptual overlap with the “motivation” construct domain.

Motivation and cognitive ability may have been moderately correlated because individuals who are high in general cognitive ability make more accurate expectancy assessments (that one can achieve a level of performance) and instrumentality assessments (that performance will lead to a desired reward), thus facilitating future motivation to perform (Vroom, 1964). The moderate correction between general cognitive ability and personality was unexpected. One's general cognitive ability may be viewed as part of an individual's overall persona or personality.

In sum, these correlations suggest moderate support for convergent and divergent validities of these scales consistent with theory-based expectations relating to the big five personality factors (Barrick & Mount, 1991) and expectancy theory (Vroom, 1964).

**Research Question 3.** The last construct validity research question addressed whether items sorted into ecology model-based construct domains demonstrated higher criterion-related validity than items not considered theory-based Q-sorters. Results suggested theory-based items did not outperform
non-theory-based items in terms of criterion-related validity. This finding runs counter to the few studies that have examined this issue (Quaintance, 1981; Redmond & Nickels, 1989; Williams, 1961). This finding must be considered tentative due to the small number of items labeled “non-theory.”

Theory-based biodata items may not have outperformed the non-theory-based items because some “constructs” probably guided original item development, though not ecology model constructs. Items labeled non-theory in terms of the ecology model could have tapped meaningful construct domains not found in the ecology model. A better test of this research question would have been to develop a set of items randomly and a set of items explicitly based on the ecology model construct domains.

**Summary of construct validity findings and conclusions.** Three sets of results permitted inferences to be drawn on the validity of latent ecology model construct domains. First, judges agreed on Q-sort classifications for the majority of biodata items relative to the ecology model construct domains, thus supporting an initial inference of content validity. Second, meaningful criterion validities were found for items falling in all but one of the five ecology model construct domains. Additionally, correlations between ecology constructs suggested moderate support for convergent and divergent validities of these scales, consistent with theory-based expectations. Third, confirmatory factor analysis results failed to support the ecology model as originally conceived (Mumford & Stokes, 1992; Mumford, Stokes, & Owens, 1990).
Subsequent iterations of exploratory factor analyses permitted interpretation of item parcel content and original confirmatory factor analysis modification indices. This Iterative Rational/Empirical approach yielded an interpretable latent three factor solution. These post hoc analyses produced a factor structure demonstrating good fit to the data when ecology model constructs were grouped according to developmental time windows. Interpretation of this 3-factor time windows solution constitutes a modification to the ecology model that may enhance future efforts to test and elaborate ecology model predictions. Processes underlying the paths in the ecology model need to be examined. The time windows perspective offers an extension to the ecology model by suggesting possible processes by which life events influence performance.

**Biodata Criterion Validity**

The second purpose of this study was to examine biodata criterion-related validity. First, simple criterion validities for biodata and general cognitive ability instruments were calculated. Second, biodata incremental validity relative to a general cognitive ability measure was examined. Interestingly, a literature review found only seven studies administered both biodata and general cognitive ability measures in criterion-related validity designs, and none examined incremental validity of either predictor. Biodata items Q-sorted as tapping general cognitive ability items were also
incrementally compared to the general cognitive ability measure. Research questions 4 through 8 specifically address criterion-related validity issues.

**Research Questions 4 and 5.** These research questions examined whether the biodata instrument and general cognitive ability test were valid predictors of job performance, respectively. The biodata instrument yielded a cross-validity of $r = .363$ (corrected for indirect range restriction, $r_c = .427$). The uncorrected cross-validity was within the range of previous findings of biodata criterion-related validity of $.30 - .40$ (Asher, 1972; Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). The general cognitive ability test yielded a correlation of $.166$ (corrected for direct range restriction $r_c = .395$). The general cognitive ability measure criterion-related validity finding is within the typical average range of corrected criterion-related validities found for general cognitive ability ($r_c = .30 - .40$; Mitchell, 1996). These correlations suggest both predictors are individually valid predictors of criterion performance.

**Research Question 6.** This research question addressed the relative contributions of biodata and general cognitive ability. Hierarchical multiple regressions (using correlation matrices corrected for range restriction as input) suggested biodata yielded more incremental predictive ability to the general cognitive ability measure ($\Delta R = .113$) than general cognitive ability added to biodata ($\Delta R = .071$). This finding could be explained by the fact that biodata has been posited to capture many different constructs which play a role in
performance (Mumford & Stokes, 1992; Mumford, Stokes, & Owens, 1990). General cognitive ability devices, conversely, are designed to measure narrowly targeted constructs. While general cognitive ability plays a role in performance, it is not the only factor (Campbell, Dunnette, Lawler, & Weick, 1970). Note significance testing could not be performed on these analyses because corrected correlations were used as input to the regression analysis (Phil Bobko, personal communication, January 1999). However, sample sizes were so large (ranging from 686 to 768) that all incremental validities calculated were surely non-zero in the population.

**Research Question 7.** Incremental validities were further examined by looking at scales created within the biodata instrument. The biodata instrument was subdivided by Q-sort results into general cognitive ability-sorted ("g-loaded") items and non-general cognitive ability-sorted ("non-g-loaded") items. Research Question 7 addressed whether "g-loaded" and "non-g-loaded" biodata items were differentially related to the general cognitive ability measure. It was expected that g-loaded items should correlate more with a general cognitive ability measure than non-g loaded items. This expectation was confirmed, as the g-loaded and non-g loaded items correlated .139 ($r_c = .345$) and .037 ($r_c = .100$) with general cognitive ability measure, respectively. The correlation between g-loaded biodata items and the general cognitive ability measure was not as high as expected. This could be explained by the fact that many biodata items sorted as tapping general
cognitive ability were probably capturing other constructs as well. For example, Q-sorters placed all items regarding high school and college grades into the general cognitive ability construct pile, though these items may also tap motivation as a secondary construct.

Criterion-related validity analyses were also performed on high school, college, and job biodata scales emerging from exploratory factor analyses. Interestingly, the farther the scale was temporally from the criterion, the higher the criterion-related validity achieved ($r_c = .32, .19, \text{ and } .17$, respectively). Job scale biodata items should have had the closest one-to-one relationship with job performance of the three scales, but this was not reflected by higher criterion validity. Education-based scales (high school and college) may have outperformed the job scale because academic performance may be more objectively measured, regularly assessed, and, quantifiable, allowing more accurate performance assessments then items measuring early career experiences. In contrast, individuals typically receive assessments of their performance on the job only once or twice a year. The performance evaluations that employees receive are often contaminated by biases and subjectivity (Bernardin & Beatty, 1984; Cardy & Dobbins, 1994; Latham & Wexley, 1981).

Research Question 8. The final criterion-related validity research question examined whether "g-loaded" biodata items, "non-g-loaded" biodata
items, and the general cognitive ability test exhibit incremental validity relative to one another.

Contrary to expectations, the non-g biodata items added less incremental prediction than g-loaded biodata items in the presence of the general cognitive ability measure ($\Delta R = .073$ and .106, respectively). The non-g biodata items were expected to add more to prediction when added to the general cognitive ability measure because they tapped constructs other than g. However, the relative difference in incremental criterion-related validity between the general cognitive ability measure and the g-loaded and non-g loaded biodata items was small (.086 and .032, respectively). Further, the cognitive ability test had higher incremental validity than high school, college, and job scales.

While the general cognitive ability test outperformed the biodata sub-scales, it did not outperform the entire biodata instrument. By dividing the biodata instrument into sub-scales, a smaller number of items were available for prediction in any one scale. With the smaller number of items per scale, lower criterion validities for the biodata scales were expected.

**Summary of criterion-related validity findings and conclusions.**
The biodata scale (including all biodata items) outperformed the general cognitive ability test both individually and incrementally (both before and after correcting for the effect of range restriction due to selection on g). This finding could be explained by the fact that biodata captures multiple constructs as per
the ecology model while tests of general cognitive ability are specifically designed to measure the more targeted construct of general cognitive ability.

When the biodata instrument was analyzed in terms of g versus non-g loaded biodata items, the general cognitive ability test outperformed the scales in terms of simple and incremental criterion validity. The finding that g biodata items added more incremental validity to g relative to non-g biodata items was unexpected. The relative difference in the incremental validities was rather small, thus tempering any conclusions regarding the incremental validity assessment of biodata item content (g versus non-g biodata items) compared to the general cognitive ability test. When the biodata instrument was investigated in terms of high school, college, and job domains, the general cognitive ability test outperformed the scales. An unexpected finding was that the farther these sub-scales were temporally from the criterion, the higher criterion-related validity achieved.

**Biodata Adverse Impact**

The final purpose of this study was to examine biodata adverse impact. Specifically, the influence of excluding response options demonstrating adverse impact from a biodata empirical key on subgroup standardized mean difference and overall biodata criterion-related validity was investigated. Additionally, adverse impact of the biodata predictor scale (both including and excluding response options displaying adverse impact) and a general cognitive
ability measure were compared. Research questions 9 and 10 specifically addressed adverse impact issues.

**Research Question 9.** This research question addressed whether there was a significant change in criterion-related validity and mean subgroup differences when adverse impact response options were removed from the empirical key. Interestingly, results indicated a substantial number of response options (44%) did not pass the four-fifths rule. This finding runs counter to the literature suggesting biodata does not adversely impact subgroups. However, the prior literature generally speaks only to the level of adverse impact found when using an overall biodata score. This study examined both response option and scale score levels of analysis to get a better understanding of the effect of adverse impact. Removal of adverse impact response options caused standardized mean difference between racial subgroup biodata scores to decrease by two-thirds. This suggests that attending to adverse impact at the response option level provides great utility for decreasing overall adverse impact at the scale score level.

Importantly, the decrease in adverse impact was not accompanied by a comparable decrement in biodata criterion validity. A minimal, non-significant decrement of .021 in biodata predictive validity occurred when adverse impact response options were removed. Adverse impact response options may lower criterion validity relative to non-adverse impact response options, though post hoc analyses suggest this was not the case in this data set. The average
criterion validity of adverse impact response options and non-adverse impact response options was .026 and .025, respectively.

Another possible explanation could be that adverse impact response options were chosen with less frequency on average across all groups compared to non-adverse impact response options. If fewer people chose those response options displaying adverse impact, these response options will necessarily have less of an effect on biodata score. Post hoc analyses showed this may partially explain current findings. The average response frequency of adverse impact response options and non-adverse impact response options was 139 and 319 respondents in the cross validation sample, respectively.

Research Question 10. A second assessment of biodata adverse impact involved more traditional assessments of scale scores. The final research question of this study addressed whether cognitive ability and biodata measures differed significantly in terms of adverse impact. First, both biodata scales (including and excluding adverse impact response options) and the general cognitive ability measure were tested for fairness using the Cleary model of test bias (Cleary, 1968). Second, a mock selection using arbitrary cut points (20th, 40th, 60th, and 80th percentiles) in the sample was performed to examine adverse impact of the selection devices.

According to Maxwell and Arvey (1993), tests with the highest criterion validity within the universe of fair tests will have the lowest adverse impact. All
three predictors exhibited test fairness using the Cleary (1968) model. The biodata instrument had higher criterion validity ($r_c = .427$) than the non-adverse impact biodata scale ($r_c = .410$) and the general cognitive ability test ($r_c = .395$). Comparison of the three predictors (biodata scale with and without adverse impact and general cognitive ability) revealed the rates at which subgroups were selected differed in a way that was consistent with Maxwell and Arvey's (1993) proof.

Maxwell and Arvey's proof applies to independent tests. The biodata scale scores with and without adverse impact response options are clearly not independent. Hence, results are consistent with Maxwell and Arvey's proof because the independent tests (entire biodata scale v. general cognitive ability test and the non-adverse impact biodata scale v. general cognitive ability test) demonstrated that instruments with the highest validity indeed had the least adverse impact. Comparison of the biodata sub-scales cannot be interpreted using the Maxwell and Arvey proof because they are not independent tests.

The biodata scale without adverse impact response options passed the four-fifths test at all cut scores, the biodata scale including all response options passed the four fifths test at the lower cut scores (20th and 40th percentiles), and the general cognitive ability test exhibited adverse impact at every cut-score. These results are also consistent with previous findings that, on average, biodata predictor scales tend to display less adverse impact than tests of cognitive ability (Reilly & Chao, 1982; Reilly & Warech, 1990).
Interestingly, relative differences in criterion validities were rather small among the three predictors (.395 - .427), yet the difference in adverse impact among the predictors varied substantially.

**Summary of adverse impact findings and conclusions.** A number of interesting findings were generated from the adverse impact analysis. First, when response options displaying adverse impact were removed from the scoring key, the standardized mean subgroup difference on the biodata instrument decreased by 66% yet criterion-related validity decreased by only 5.7%. Both biodata scales (all items versus non-adverse impact items) and the cognitive ability test passed the Cleary model of test fairness and exhibited similar corrected criterion validities. When the three predictors were used in a mock selection with percentile selection cut scores, the biodata scale without adverse impact items outperformed the other predictors in terms of adverse impact.

**Future Research**

**Biodata theory.** The American Psychological Association's (APA) Task Force on Statistical Inference recently suggested there is a need for more theory generating studies relative to theory confirming studies. The Task Force suggested researchers are "forced into the premature formulation of theoretical models in order to have their work funded or published" (p. 2, APA Task Force, 1996). Additionally, researchers need to be more receptive to well conducted exploratory research to enhance the quality and utility of future
theory generation and assessment. This recent recommendation serves to emphasize the importance of exploratory research to improve and refine theory.

This study was originally undertaken as a confirmatory study of the ecology model on an organizational data set. The fact that the ecology model did not receive strong support in the confirmatory analysis indicated further exploratory research on biodata theory is needed. The most interesting finding regarding biodata theory came from the follow-up exploratory analysis undertaken as a result of poor initial ecology model fit. The exploratory analysis found support for the ecology model when constructs were grouped based on a time windows perspective (Rovee-Collier, 1995) rather than simply grouping the items by construct. This approach, labeled an Iterative Rational/Empirical approach, represents a cycle of exploratory and confirmatory analysis aimed at continuous theory improvement. Additional research is needed to determine whether a "time windows ecology model" is a more accurate conceptualization processes underlying biodata prediction.

Future research is needed to find other highly developmental predictive and theoretically important time periods (such as later life or career events) not captured in this instrument. After additional exploratory research is conducted, confirmatory research should be undertaken using biodata instruments based on the time windows-ecology model framework to test the model extension proposed by this research.
Biodata criterion validity. It is recommended that the questions addressed in this study using this item pool, job, and set of applicants be replicated in other item pools, jobs, and applicant pools to see if results generalize to other populations.

Biodata adverse impact. Future research should determine the decrement in validity of item versus response option deletion. In this study, 92% of all the biodata items had at least one response option fail the four-fifths criterion. This suggests response option level modifications may be the best means of diminishing adverse impact while retaining predictive power.

Practical Implications

Biodata criterion validity. This research has practical implications for biodata and general cognitive ability measures. Both predictors yielded incremental criterion validity though biodata exhibited greater simple and incremental criterion-related validity. Biodata may be used in replacement of or in concert with general cognitive ability measures. Using biodata, organizations may reap the benefits of biodata's traditionally high criterion-related validity and low adverse impact relative to measures of general cognitive ability. Organizations screening applicants based solely on general cognitive ability test scores are likely to incur a slight performance decrement relative to organizations using biodata and a severe performance decrement relative to those organizations using both.
**Biodata adverse impact.** This study offered a practical implication for minimizing biodata adverse impact while also minimizing decrements in predictive ability. Simple cross-tabulation of racial subgroup response option frequency in key development samples can determine which response options demonstrate differential response frequencies for protected subgroups. This analysis should be routinely applied in biodata response option-based empirical key. Elimination of response options demonstrating adverse impact from the scoring key can yield substantially lower adverse impact and high predictive validity. Organizations interested in selecting a diverse group of high performing individuals should not be basing selection decisions solely on general cognitive ability tests scores. Reliance on general cognitive ability alone will only yield lower criterion-related validity and higher adverse impact.

**Study Limitations**

This study suffered from two primary limitations in drawing inferences for theory and practice. These limitation focus on 1) the post hoc nature of the test of biodata theory, and 2) sample range restriction due to selection on the test of general cognitive ability. The first limitation was that post hoc tests of the ecology model were performed on an existing biodata instrument, fitting the model to items that were made available to the author. This might explain why confirmatory factor analysis results did not support the ecology model. Ideally, the items would have been developed a priori specifically to test the ecology model.
This study does not constitute a strong test of the ecology model because of its post hoc nature, though it does provide an important initial test of the theory on a non-student sample. Research leading to the development of the ecology model was conducted exclusively on student samples, suggesting a threat to external validity (Cook & Campbell, 1979). The current research addresses the external validity issue by applying the model in an organizational setting.

Additionally, there may be other constructs that biodata captures that were not included in the ecology model, thus leading one to question the study's internal validity. However, the intent of the current study was to perform a confirmatory test of the ecology model as it is currently defined. Exploratory analyses did suggest modifications to the ecology model using the concept of time windows may help to understand underlying processes behind the constructs proposed by the ecology model.

A second limitation was that the data were very range restricted due to selection on the general cognitive ability measure. This limitation was dealt with by using statistically corrected correlation matrices for simple and multiple regression analyses. The problem of range restricted data is not unique to this study. Most personnel selection studies use predictive validation with selection designs, where only those applicants who were selected have criterion measures available (e.g., Russell & Dean, 1995; Schmitt, Gooding, Noe, & Kirsch, 1984). However, a strength of this data was the large sample
upon which to draw conclusions. Over 10,000 individual's predictor and criterion data were available for analysis as well as data on the population of 206,000 applicants who took the general cognitive ability as an initial selection screen. These data permitted estimation of the unrestricted standard deviation in the general cognitive ability test scores in the entire applicant population enabled accurate corrections for range restriction.

Conclusion

This study offers contributions for both biodata theory and practice. A major theoretical contribution of this study is the finding that the ecology model's five construct domains may be best conceived within developmental time windows. For example, items tapping the interpersonal skill domain in high school seem to capture a meaningfully different construct than items tapping interpersonal skill at early career entry. It remains to be seen whether evidence will support a model containing five independent latent constructs within periods of change and development. Regardless, results suggest continued programmatic research holds great promise for developing a strong theory of biodata.

This research has practical implications for biodata and general cognitive ability measures. Both yielded incremental validity, though biodata exhibited greater simple and incremental validity. Additionally, biodata's already low adverse impact can be further improved by simply examining adverse impact at the response option level and removing those response
options exhibiting adverse impact. Removal of response options exhibiting adverse impact resulted in a minimal decrement in biodata predictive ability and a substantial decrease (two-thirds) in standardized mean difference between black and white biodata scores.

In conclusion, this research provided much needed evidence examining biodata theory in an organizational setting and offered practical implications for biodata applications. Results of this study suggested further refinement of existing theory in terms of conceptualizing theory using the time windows approach. In terms of the practical application of biodata, the selection technology appears to have the ability to help organizations meet two important objectives simultaneously – selecting the best job candidates while at the same time encouraging workforce diversity.
REFERENCES


Cascio, W. F. (1975). Accuracy of verifiable biographical information blank...


Greenwald, A. G. (1975). On the inconclusiveness of crucial cognitive tests of


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


Hillsdale, NJ.


predictors of occupational performance. Paper invited by the National Commission on Testing and Public Policy, Washington, DC.


APPENDIX

LETTER OF PERMISSION

Michelle Dean, Ph.D.
3232 N. Locust Street, Apt. 626
Denton, TX 76207

Dear Ms. Dean,

In support of agency research objectives on alternative selection measures for the air traffic control specialist occupation, you are granted permission to use archival biodemographic, cognitive aptitude test, and training performance data and measures in your dissertation on the validity and fairness of empirical keying of biodata instruments. The data are provided for research purposes only, and may not be used for any commercial purpose. You agree to acknowledge the FAA as the source for your research data, and provide a bound copy of your doctoral dissertation to the FAA.

Edna Fiedler, Ph.D.
Manager, Training and Organizational Research Laboratory

March 29, 1999

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

Michelle A. Dean received a bachelor of science degree in Management and a master of Business Administration degree from Louisiana State University. She has co-authored papers in the *Academy of Management Journal* and *Research in Personnel/Human Resource Management*. She has presented her research at the annual meetings of the Academy of Management and the Society for Industrial and Organizational Psychology. Her research interests include personnel selection, performance prediction, and research methods in human resource management. She completed the requirements for the doctor of philosophy degree in Business Administration in 1999.
Candidate: Michelle Ann Dean

Major Field: Business Administration (Management)

Title of Dissertation: On Biodata Construct Validity, Criterion-Related Validity, and Adverse Impact

Approved:

[Signatures]

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination:

March 5, 1999