Technical Report 939

Capturing Temperament Constructs With Objective Biodata

Fred A. Mael and Amy C. Schwartz
U.S. Army Research Institute

November 1991

United States Army Research Institute
for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

MICHAEL D. SHALER
COL, AR
Commanding

Technical review by

Henry Busciglio
Robert Kilcullen

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 939	5. MONITORING ORGANIZATION REPORT NUMBER(S) --
6a. NAME OF PERFORMING ORGANIZATION U.S. Army Research Institute 6b. OFFICE SYMBOL (If applicable) PERI-RS	7a. NAME OF MONITORING ORGANIZATION --
6c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	7b. ADDRESS (City, State, and ZIP Code) --
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-R	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER --
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62785A; PROJECT NO. 791; TASK NO. 2211; WORK UNIT ACCESSION NO. H1
11. TITLE (Include Security Classification) Capturing Temperament Constructs With Objective Biodata	
12. PERSONAL AUTHOR(S) Mael, Fred A.; and Schwartz, Amy C.	
13a. TYPE OF REPORT Interim	13b. TIME COVERED FROM 89/11 TO 91/07
14. DATE OF REPORT (Year, Month, Day) 1991, November	15. PAGE COUNT
16. SUPPLEMENTARY NOTATION -- fy back	

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Biographical data (biodata), ~~USA~~ Military Academy, Temperament, * Biographies, * Aptitude tests

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The purpose of this research is to develop biographical data (biodata) indicators of attrition and leadership potential and performance that can measure relevant temperament constructs suitable for an admissions package at the U.S. Military Academy (USMA). The Army temperament measure Assessment of Background and Life Experience (ABLE) and a 73-item biodata instrument were administered to 1,325 incoming USMA cadets in July 1990. The biodata items were used to produce analogs to five ABLE scales. The relationship of the ABLE scales and their biodata analogs to three criteria (attrition from 6-week summer training, leadership ratings from the 6-week period, and leadership ratings from fall semester) and the incremental contributions of ABLE and biodata analogs beyond the current USMA Whole Candidate Score (WCS) were evaluated. Results demonstrated that the biodata scales showed strong relationships to their equivalent ABLE scales and comparable relationships to the criteria in 13 of 15 comparisons. Also, for each criterion, either overall ABLE or the biodata equivalent added incremental validity over and above the WCS. Five of the six

(Continued)

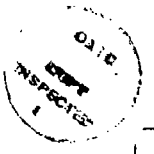
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael G. Rumsey	22b. TELEPHONE (Include Area Code) (703) 274-8275 22c. OFFICE SYMBOL PERI-RS

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

ARI Technical Report 939

19. ABSTRACT (Continued)

biodata scales had significantly smaller correlations with a social desirability scale than the equivalent ABLE scale. The results demonstrate the potential role of a biodata measure in an admissions package as a versatile indicator of attrition and leadership ability.



Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Technical Report 939

Capturing Temperament Constructs With Objective Biodata

Fred A. Mael and Amy C. Schwartz
U.S. Army Research Institute

Selection and Classification Technical Area
Michael G. Rumsey, Chief

Manpower and Personnel Research Laboratory
Zita M. Simutis, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

November 1991

Army Project Number
2Q162785A791

Manpower, Personnel, and Training

FOREWORD

A primary objective of research task 2211H1 of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is to provide enhancements to selection through development and refinement of new measures.

The focus of this research is to develop biodata indicators of attrition from training and leadership potential and performance that will measure relevant temperament constructs and be suitable for use in an admissions package at the U.S. Military Academy (USMA). The results of this phase of the research indicate that biodata scales can be used to provide indexes of attrition from training and leadership performance during a cadet's first 6 months at USMA. In addition, the biodata measures demonstrate properties in cadets that make them more suitable for admissions than their temperament counterparts. Moreover, the temperament and biodata measures add incremental validity over and above that of measures currently used for admissions to USMA.

This research is the result of a collaborative effort between the Office of Institutional Research (OIR) at USMA and ARI initiated in November 1989. The commander and researchers at OIR have been apprised of research results on a continuous basis. Follow-up research will include cross-validation of results and additional measures of performance from subsequent stages of the cadets' tenure at USMA and in the officer corps.

EDGAR M. JOHNSON
Technical Director

ACKNOWLEDGMENTS

We would like to acknowledge the help of William Burke of the Office of Institutional Research (OIR) at the U.S. Military Academy (USMA). Dr. Burke provided and continues to provide sage advice and direction on all aspects of this research effort. Jane Richards and Jackie Pittard also provided invaluable technical assistance. Additional thanks must go to Colonel Patrick Toffler for his support of our efforts. We hope that our findings will be useful for OIR and USMA in the future.

Leonard White's extensive experience with the Army ABLE, his methodological acumen, and his support were drawn upon regularly at all stages of this research. Frank Erwin, Michael Mumford, and Richard Reilly were among a number of researchers and practitioners who freely shared their expertise in the field of biodata. Michael Rumsey, Henry Busciglio, and Robert Kilcullen provided initial reviews of the paper. Their many useful suggestions have been incorporated.

CAPTURING TEMPERAMENT CONSTRUCTS WITH OBJECTIVE BIODATA

EXECUTIVE SUMMARY

Requirement:

The purpose of this research is to develop biodata indicators of attrition from training and leadership potential and performance that will measure relevant temperament constructs, yet still be potentially suitable for use in an admissions package at the U.S. Military Academy (USMA).

Procedure:

The Army temperament measure Assessment of Background and Life Experience (ABLE) and a 73-item biodata instrument developed for this research were administered to 1,325 members of the USMA Class of 1994. Criterion measures were attrition from the 6-week preliminary summer training period, leadership ratings from that summer period, and leadership ratings from the fall semester. The biodata items were coded in order to produce analogs to the five ABLE scales in the research. The relationship of each ABLE scale and its biodata analog to each of the three criteria, as well as the incremental contributions of the total ABLE and its biodata analog over and above that of the currently used USMA Whole Candidate Score (WCS), were evaluated. In addition, the relative contribution of each component ABLE scale as an indicator for each criterion was assessed. Finally, the susceptibility of both the ABLE and biodata scales to socially desirable responding was investigated.

Findings:

The biodata scales showed strong relationships to their equivalent ABLE scales and smaller relationships to the other ABLE scales. When compared with the ABLE scales regarding their relationship to the criteria, the biodata measures demonstrated comparable validities in 13 of 15 cases. Further, for each criterion, either overall ABLE or the biodata equivalent added incremental validity over and above the WCS. Four of the five individual biodata scales, as well as the overall biodata scale, had significantly smaller correlations with a social desirability scale than the equivalent ABLE scale.

Utilization of Findings:

The results of this research can be used to develop an indicator of attrition and leadership potential that will enhance the USMA admissions package. The research also refined the methodology for developing biodata analogs to temperament measures. This methodology will prove useful in ongoing investigations of the feasibility for using these measures in officer and enlisted selection.

CONTENTS

Page

INTRODUCTION 1

Research Problem: The Keying Dilemma 1
The Research Program at USMA 2
Current USMA Admissions Procedures 3
The Army's ABLE 3
What Makes Biodata Biodata? 5
Defining Biodata 6
Biodata Item Attributes 7
Summary 10
The USMA Research Effort 10

METHOD 11

Sample 11
Questionnaire Development 11
Keying Procedures and Strategies 13
Keying to ABLE Scales 14

RESULTS 15

ABLE and Bioabsum 15
Interrelationships Among Predictors 17
Relationships to Six-Week Attrition 18
Relationships to Six-Week Leadership Ratings 19
Relationships to Fall Semester Leadership Ratings 20
Social Desirability Analyses 21

DISCUSSION 23

REFERENCES 27

APPENDIX: EXAMPLES OF EMPIRICAL KEYING EMPLOYED IN USMA RESEARCH A-1

Page

LIST OF TABLES

Table 1. Temperament scales by construct in the Assessment of Background and Life Experience (ABLE) 4

2. A taxonomy of biodata items 8

3. Descriptive statistics and intercorrelations for ABLE scales 16

4. Intercorrelations between ABLE-keyed biodata scales and ABLE scales 16

5. Intercorrelations between ABLE, ABLE-keyed biodata (bioabsum), and current USMA predictors 17

6. Correlations of ABLE scales, biodata keyed to ABLE, and USMA attrition and leadership criteria 20

7. Correlations of ABLE scales, biodata keyed to ABLE, and USMA predictors with ABLE validity (social desirability) scales 22

CAPTURING TEMPERAMENT CONSTRUCTS WITH OBJECTIVE BIODATA

INTRODUCTION

In recent years, there has been a good deal of interest in the use of biodata for military selection, highlighted by a number of current efforts in the joint and individual service arenas (Trent, Quenette, & Pass, 1989; Watson, 1989). Reviews of selection measures have found biodata validity coefficients to be impressive compared with other measures (Asher & Sciarrino, 1974; Ghiselli, 1966; Reilly & Chao, 1982). Recent research indicates that biodata validities may be more stable over time and more generalizable across organizations than previously thought (Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990). Nevertheless, certain concerns involving the use of biodata in applied settings remain. This paper describes a research effort focused on dealing with these concerns, conducted by members of the U.S. Army Research Institute for the Behavioral and Social Sciences in conjunction with the Office of Institutional Research at the U.S. Military Academy at West Point.

Research Problem: The Keying Dilemma

Currently, two methods are most commonly used for keying biodata, that is, determining the numerical value (weight) to be assigned to each response alternative within an item (Mumford & Owens, 1987). The first approach, empirical keying, was the sole method used in early biodata research and continues to be used by many practitioners. With empirical keying, weights are assigned to each alternative based on its mean score on the criterion being used. For example, if the criterion is leader ratings, the value on an item which has the highest average rating is assigned the highest score. The same is done for each alternative, so that the continuum of values within the item is arranged to reflect scores on the criterion. Purely empirical keying is highly sensitive to sample characteristics and can thereby lead to an optimal correlation with the criterion. However, when the key is cross-validated, the regression coefficient often is much smaller than that of the initial derivation sample, a phenomenon referred to as shrinkage. Moreover, the method has been termed "dustbowl empiricism" by critics for being atheoretical and failing to advance understanding of the underlying antecedents of successful performance (Dunnette, 1962; Pace & Schoenfeldt, 1977).

Some researchers have instead championed a rational approach to biodata, in which item alternatives are assigned a priori values based on a presumed relationship of the item to a specific, unitary construct (Mitchell & Klimoski, 1982). Thus, the rational approach is usually an attempt to measure temperament or other constructs with biodata-like items. Adherence to this strategy leads to a preference for items that can be clearly related to only a single construct and then combined into homogeneous scales. Advocates of rational biodata development claim that their method, which uses predetermined values for item responses, will reduce shrinkage because it is not fitted to

sample-specific idiosyncracies. A possible problem with this approach, however, is that responses to complex, heterogeneous behaviors would also have to be coded in terms of single constructs, even if they were really a function of multiple influences. Also, by making a priori decisions about item directionality across any or all criteria, the possibility that a certain behavior will be beneficial for some outcomes and counterproductive for others is generally downplayed.

In recent years, another issue has surfaced. Many researchers have expressed concern about the possibility of socially desirable responding and faking on self-report measures, notably temperament measures (Crowne & Marlowe, 1960; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Paulhus, 1984). While the same concern has been expressed about biodata, one proposed solution has been to limit biodata to objective and verifiable items. This presents a problem for those advocating rational keying, in that objective and verifiable actions tend to be heterogenous (determined by multiple causes) and therefore difficult to attribute to a single temperament. Conversely, adherence to this strategy would eliminate the use of subjective, homogenous items. Because researchers using the rational approach do not limit themselves to objective, historical and/or verifiable items, their measures are often indistinguishable from temperament scales, and may be more fakable than empirically keyed biodata.

In this research effort, an attempt was made to gain the conceptual benefits of rational methods, while gaining the less fakable properties associated with objective, verifiable biodata. Specifically, an attempt was made to key verifiable biodata directly to temperament scales, and then use those scales rationally with multiple criteria. Though no attempt was made to assign items exclusively to a single construct, the goal was to determine if biodata scales could be utilized to parallel individual temperament scales. A secondary concern, if this method constitutes an optimal use of objective biodata items compared with some form of direct empirical keying, is to be covered in subsequent research.

The Research Program at USMA

Leadership research has been an abiding interest of the U.S. Military Academy (USMA) at West Point for many years (Page, 1934). A current example of this emphasis is the Leadership Development Project, an ongoing research effort, approved in March, 1988, and directed by the USMA Office of Institutional Research (OIR). The stated goals of the project are to a) improve measurement of candidate leader potential, b) improve measurement of cadet leadership performance, and c) measure contributions of USMA graduates to the common defense. In late 1989, OIR and the U.S. Army Research Institute (ARI) decided to jointly pursue their mutual interest in working on the first of these goals.

USMA uses a three-pronged approach to selection, attempting to find candidates who will excel academically, physically, and militarily. As officer military excellence is

largely defined as the ability to lead others, measurement of leadership potential and performance is a priority. OIR/USMA felt that although candidate academic and physical capabilities were adequately measured, improved measurement of leadership potential was possible. In addition, while SAT scores have had a strong, demonstrated relationship to academic attrition, interest was expressed in finding better indicators of nonacademic attrition.

Current USMA Admissions Procedures

Currently, an important indicator of candidate potential used in the admissions decision at West Point is based 60% on an applicant's standardized test scores (i.e., SAT, ACT) and graduating rank in high school; 30% on the Leadership Potential Score (LPS), derived from the School Official Evaluation (an evaluation form filled out by high school instructors), and the Candidate Activities Record (CAR), a checklist of extracurricular activities and varsity sports; and 10% on scores on the Physical Aptitude Examination (PAE). This information is combined in a weighted composite known as the Whole Candidate Score (WCS). The information on the CAR is similar to that on a biodata instrument. However, scoring keys were based on content validity judgments of USMA personnel rather than on criterion-related validity judgements. This raised the possibility that an alternative approach might yield higher relationships with relevant outcome measures.

In conjunction with the Leadership Development Project, OIR explored approaches toward improving USMA's selection procedures with two goals in mind. The first was to attempt to capture motivational indicators of leadership performance, with the possible goal of including these measures in future admissions packages. The second was to explore empirical methods of scoring new and existing inventories of previous behaviors and experiences. Accordingly, OIR/USMA suggested that ARI administer a biodata questionnaire, as well as ABLE, an Army temperament measure, in order to accomplish these goals. The next section describes ABLE, the Army's temperament test, and its role in the current project. The following section defines what a biodata measure is, especially in contradistinction to a temperament measure, and describes the guidelines considered in the development of the USMA biodata measure. Finally, the specific approach and goals of this research are outlined.

The Army's ABLE

The Assessment of Background and Life Experiences (ABLE), is a temperament measure developed and validated by the U.S. Army Research Institute as part of a long-term research program called Project A, which was designed to revalidate the Armed Services Vocational aptitude battery (ASVAB) and design supplementary tests measuring additional constructs. ABLE was included in Project A to capture the motivational element of performance ("will do"), as opposed to the ability ("can do") element. ABLE

is also under consideration in the Joint Services arena as part of a measure of adaptability to military life.

Scale development involved reviewing 12 major personality inventories and then reducing the number of temperament dimensions by eliminating redundancy and focusing on predictors of job performance (Hough et al., 1990). In its complete form, ABLE consists of 10 scales measuring 5 constructs (see Table 1). In addition, validity scales, which indicate whether or not the respondents answers reflect faking or social desirability distortion, were included. Nearly 50,000 soldiers in 21 Military Occupational Specialties (MOS) were tested and their ABLE scores were used to predict NCO leadership potential, disciplinary problems, and attrition.

Table 1

Temperament Scales by Construct in the Assessment of Background and Life Experience (ABLE)

Construct	Scale
Stress Tolerance	Emotional Stability
Dependability	Nondelinquency Traditional Values Conscientiousness
Achievement/Leadership	Work Orientation Self-Esteem Dominance Energy Level
Physical Condition	Physical Condition
Locus of Control	Internal Control
Agreeableness/Likability	Cooperativeness
Response Validity Scales	Non-Random Response Social Desirability

Previous findings indicate that ABLE predicts enlisted attrition, effort and leadership, and personal discipline (Hough et al., 1990). In Project A research, lower scores on ABLE were significantly related to greater rates of attrition, with the

relationship most pronounced among those scoring low on ABLE (White, Nord & Mael, 1990). ABLE was also related to probability of graduation of USMA graduates and other trainees at the Ranger school course. In addition, the Achievement construct of ABLE was found to be a significant predictor of effort and leadership (for the Work Orientation scale, uncorrected r = .23). Other scales which significantly predicted effort and leadership included Dominance, Energy Level, and Emotional Stability. Finally, the ABLE Dependability construct significantly predicted discipline problems among enlisted soldiers, with (uncorrected) validities ranging from .23 to .29 for the three Dependability scales. ABLE was thus seen as an attractive measure of adaptability because it specifically addressed dominance and leadership proclivities, and because of its documented relationship to prediction of attrition, indiscipline, and leader potential among NCOs.

However, ABLE has potential drawbacks for use in an enlistment or admissions package. One is the fear of extensive faking and socially desirable responses. ABLE is a relatively transparent test, with no attempt to obscure desirable responses, and with virtually all items arranged in a linear continuum of desirability. In a previous administration with enlisted soldiers, faking has not contaminated ABLE's validity (Hough, et al., 1990). However, the fear of faking would be increased in an admissions situation, where the instrument is often taken at home under the tutelage of parents and other advisors. A second concern was that some ABLE items concerned somewhat intrusive and "psychological" topics, such as physical symptoms, fears, anxieties, and feelings of depression and failure. USMA researchers felt that these types of items could be resented, thus driving away capable candidates.

Therefore, the researchers sought to determine if ABLE constructs could be measured with more palatable biodata items. Given that a biodata measure was sought specifically because of the qualities distinguishing it from temperament measures, it became crucial to define the unique characteristics of biodata, as well as how they differ from temperament measures. The guidelines which emerged from this effort are described next.

What Makes Biodata Biodata?

There is considerable controversy regarding the criteria for specifying the domain and attributes of biodata items (Asher, 1972; Gandy, Outerbridge, Sharf, & Dye, 1989; Henry, 1965; Stricker, 1987). In addition, while some have attempted to differentiate biodata items from temperament, attitude, or interest items (Guthrie, 1944; Mumford & Owens, 1987), in practice, many items termed "biodata" are indistinguishable from self-report temperament items (Crosby, 1990). It is not uncommon to find items about internal states, opinions, and reactions to hypothetical situations included in biodata measures. The result has been a continued blurring of what constitutes biodata.

The confusion is especially problematic in light of claims that biodata scales are more resistant to social desirability distortion (Telenson et al., 1983) and generally achieve higher validities (Asher, 1972; Reilly & Chao, 1982) than temperament measures. However, this may be true only of certain types of biodata, such as verifiable items. It is therefore worthwhile to enumerate the attributes that have been used to define biodata and differentiate it from other self-report measures.

Defining Biodata

Biodata items attempt to measure previous and current life events which have shaped the behavioral patterns, dispositions, and values of the person. Owens has stated that "one of our most basic measurement axioms holds that the best predictor of what a man will do in the future is what he has done in the past" (1976, p. 625). It is presumed that a person's outlook is affected by life experiences and that each experience has the potential to make subsequent life choices more or less desirable, palatable, or feasible. One possible reason is that the focal experience reinforces a pattern of behavior. Alternatively, the focal experience may be partly or wholly determined by earlier causal determinants- genetic, dispositional, or learned- which account for variations in both earlier and current behavior.

Moreover, every experience or series of experiences which conceivably categorizes (or stigmatizes) a person has the potential to shape that person's behavioral patterns, though each component's influence is mitigated by the effects of all other identifications. Thus, when a person associates with a team, club, school, or any other "psychological group," the person takes on (to varying degrees) the aspirations, preferences, values, and self-perceptions which are endemic to group members. Even negative categorizations (e.g., the inability to swim, ride a bike, or drive a car at the same age as classmates), or so-called "input variables" (Owens & Schoenfeldt, 1979), such as place of upbringing, size of high school, and parental occupation, could place the person in a self-perceived category with a specific profile.

From a biodata perspective, therefore, previous events and experiences are not only indications of underlying dispositions, but are themselves seen as shapers of subsequent behavior. By contrast, temperament measures primarily attempt to capture somewhat stable dispositional tendencies. Thus, the typical temperament item asks the respondent direct questions about dispositions. Alternatively, a temperament item may infer the construct from tendencies evident in narrowly focused reactions to past and current events, or from expressed responses to hypothetical and future situations. When reactions are sampled, they are seen merely as outcomes of the pre-existing temperament.

In summary, the realm of biodata is more inclusive than temperament in terms of content, in that it includes behavioral antecedents and indicators of skills, abilities, and temperaments (Mumford & Stokes, 1991). Conversely, because biodata items attempt to

measure only events and behaviors that have definitely occurred, many researchers have argued that biodata items be more restrictive in their attributes than temperament ones. The attributes fall into two categories: those that aim at increasing the accuracy of the information generated as biodata, and those that seek to limit the domain of biodata content on legal or ethical grounds. These attributes are drawn partly from an earlier typology by Asher (1972), and includes new categories mentioned previously by others (Barge, 1987; Stricker, 1987). They have been reviewed in depth by Mael (in press), and are summarized below. Examples of each attribute appear in Table 2.

Biodata Item Attributes

Historical versus hypothetical. Biodata items should pertain solely to historical events, events which have taken place, or continue to take place. This would exclude items about behavioral intentions or about presumed behavior in a hypothetical situation. This appears to be the core attribute of biodata items.

External versus Internal. Some have argued that biodata items should deal with external, though not necessarily publicly seen, actions. This would exclude items about thoughts, attitudes, opinions, and unexpressed reactions to events. Items about what one typically does in situations could be considered historical and external. While the external attribute, as well the objective and verifiable attributes mentioned below, have been ignored by a number of researchers (e.g. England, 1971; Glennon, Albright, & Owens, 1966; Russell, Mattson, Devlin & Atwater, 1990), each may be crucial to claims of greater freedom from distortion for biodata compared to temperament scales.

Objective and First-hand versus Subjective. Some who endorse the external attribute also feel that biodata should be objective recollections, requiring only the faculty of recall. Subjective interpretation of events, such as assessing if one was disappointed, angry, or depressed in a given situation, would not fit this attribute. Evaluation of one's qualities or performance relative to that of others also would be considered subjective. A corollary would be that biodata items ask only for the first-hand knowledge of the respondent, as opposed to estimation of how others (peers, parents, teachers) would evaluate one's performance or temperament, which involves an additional level of speculative subjectivity.

Discrete versus Summary Actions. Methodologically, it may be preferable to focus on discrete actions, dealing with a single, unique behavior (e.g., age when received driver's license), as opposed to summary responses (e.g. average time spent studying). Responses to summary items also require computation or estimation and increase the chance of inaccuracy. However, with a regularly performed behavior, summary recall could be more realistic and accurate than recall of a single, arbitrarily chosen instance.

Table 2:

A Taxonomy of Biodata Items

Historical How old were you when you got your first paying job?	Future or hypothetical What position do you think you will be holding in ten years? What would you do if another person screamed at you in public?
External Did you ever get fired from a job?	Internal What is your attitude toward friends who smoke marijuana?
Objective How many hours did you study for your real-estate license test?	Subjective Would you describe yourself as shy? How adventurous are you compared to your co-workers?
First-hand How punctual are you about coming to work?	Second-hand How would your teachers describe your punctuality?
Discrete At what age did you get your driver's license?	Summative How many hours do you study during an average week?
Verifiable What was your grade point average in college? Were you ever suspended from your Little League team?	Non-verifiable How many servings of fresh vegetables do you eat every day?
Controllable How many tries did it take you to pass the CPA exam?	Non-controllable How many brothers and sisters do you have?
Equal access Were you ever class president?	Non-equal access Were you captain of the football team?
Job relevant How many units of cereal did you sell during the last calendar year?	Not job relevant Are you proficient at crossword puzzles?
Non-invasive Were you on the tennis team in college?	Invasive How many young children do you have at home?

Verifiable. A verifiable item is an item that can be corroborated from an independent source. Item verifiability thus goes beyond both the external event and objective criteria. The optimal source of verification is archival data, such as school transcripts or work records. Alternatively, the testimony of knowledgeable persons, such as a teacher, employer, or coach, is also considered verification by most researchers. Asher (1972) and Stricker (1987) have advocated exclusive use of verifiable items, though others utilize or condone the use of non-verifiable items (e.g., England, 1971; Glennon, et al. 1966) and some advocate interleaving verifiable and non-verifiable items (Landy & Trumbo, 1980; Mumford & Stokes, 1991). Merely warning respondents that answers will be verified can also reduce faking (Schrader & Osburn, 1977). Verifiability should be less necessary with discrete and publicly witnessed items for which "faking good" would require conscious lying. When developing biodata, obscuring the "right" answers and using subtle items also should discourage socially desirable responses, even without the threat of verification.

Controllable and Equally Accessible. From the perspective that all life events can potentially shape and affect later behavior, there is no reason to differentiate between experiences that a person has consciously chosen to undertake and those that were components of the person's environment. Accordingly, the biodata instruments of numerous researchers include both controllable and noncontrollable items (e.g., Mumford & Stokes, 1991; Richardson, Bellows, & Henry, 1985; Russell et al., 1990). Stricker (1987), on the other hand, argues that it is unethical to evaluate people based on noncontrollable items pertaining to parental behavior, geographic background, or socioeconomic status. He also considers items dealing with skills and experiences not equally accessible to all applicants, such as tractor-driving ability or playing varsity football, to be unfair. Similarly, the developers of the Armed Services Applicant Profile (ASAP) and the Air Force's Leadership Effectiveness Assessment Profile (LEAP), two biodata measures for military use, have also attempted to delete all non-controllable items from their instrument (Trent, Quenette, & Pass, 1989; Watson, 1989).

In practice, however, strict adherence to these restrictions would lead to exclusion of most life experiences likely to be related to later behavior, as well as many items typically found on school and job application blanks. This would present an especially severe constraint when sampling applicant pools without extended job histories, such as military applicants. Because of this constraint, the LEAP researchers felt compelled to compensate with "behavioral intention" items (Watson, 1989), non-historical speculations about behaviors. Therefore, for both conceptual and practical reasons, it is argued that these two attributes need not be adhered to.

Visibly Job Relevant. Virtually all life experiences are potentially "job relevant" if they contribute to the skill base, self-efficacy, or values of the individual, even if the prospective job has no activities that are superficially analogous to the previous experience. Nevertheless, Gandy et al. (1989), citing legal constraints, feel that at least in the public sector, this type of job relevance may be insufficient. If job relevancy needs

to be narrowly defined as showing face valid job pertinence then the domain of truly relevant items would be severely limited. Moreover, paradoxically, items which fit the narrowest definition of job relevant would be the most transparent and most fakable.

Invasion of privacy. A final concern, which pertains to all self-report items, involves invasion of privacy. Many items pertaining to topics such as national origin, religious or political affiliation, or financial status, may fall afoul of Federal, state, or local privacy protection laws (Arvey, 1983; Gandy et al., 1989; Van Rijn, 1980). Genuinely intrusive questions, such as those dealing with sexual behavior, bodily functions, or specific religious and ethnic practices, are also likely to incur resistance and resentment and thereby encourage willful faking, random responding, or other behavior aimed at foiling the testers. Unfortunately, the parameters of intrusiveness and invasion of privacy have yet to be defined clearly in the literature.

Summary

The core attribute of a biodata item is that it addresses an historical event or experience. The rationale is that previous events shape the behavioral patterns, attitudes, and values of the person, and combine with individual temperaments to define the person's identity. Other attributes, though not defining biodata, may have the advantage of minimizing social desirability distortion. These include limiting items to those regarding external events, those requiring only objective, first-hand recollection, and those pertaining to verifiable events. Items involving discrete, unique events may also be preferred when appropriate. Exclusive use of controllable and equally accessible items, as well as items narrowly defined as "job relevant" should not be required unless legally mandated. While clearly intrusive items are offensive and probably counterproductive, definition of invasiveness remains unclear.

Because of concerns about faking associated with subjective items, the items used in the current research effort were all historical, external, objective, and first-person, and primarily verifiable, at least in principle. Both controllable and non-controllable items were used, and "relevance" was of necessity defined broadly. Attempts were made to avoid invasive or otherwise inflammatory items.

The USMA Research Effort

As mentioned above, OIR researchers sought to determine if temperament constructs, specifically those in the ABLE, could be measured with biodata items without loss of validity. To do this, ABLE scales deemed most appropriate for the USMA candidate pool were selected. Next, biodata items were developed which would be keyed to the appropriate ABLE scales. Because ABLE's relationship with enlisted attrition and leadership potential has been demonstrated, linking biodata to the ABLE

could determine 1)if the same relationships hold for cadets, 2) whether the ABLE constructs could be adequately measured by objective, verifiable biodata items, and 3) under these conditions, would the ABLE and the biodata only account for the same variance in attrition and leadership, or would each contribute uniquely to accounting for variance in these criteria.

Also, as opposed to other attempts at rational biodata development, the current approach takes advantage of the possibility that the behaviors or events behind each biodata item may be a result of or an antecedent of several different temperaments. For example, family birth order may be predictive of both dominance and emotional stability, while classroom performance may be related to work orientation and energy. The multidimensionality of objective life events, although problematic for the typical temperament scale, is an important feature of biodata which should be capitalized upon, rather than ignored. Conversely, perhaps keying first to ABLE constructs, and then using the predetermined key without reference to the criterion, would show greater immunity to shrinkage than that typical of empirically keyed biodata.

In summary, there were a number of important purposes for administering both ABLE and biodata at West Point. First, the feasibility of using objective and verifiable biodata items to measure temperament constructs was explored. Second, keying biodata to ABLE was examined as a quasi-rational approach which would enable the use of an empirically derived biodata measure without the shrinkage in validities often associated with criterion-keyed measures. Third, biodata and temperament analogs were compared in terms of their relationship with attrition and leadership, as well as their vulnerability to faking. Finally, the incremental contribution of both ABLE and the biodata analogs over and above that of the Whole Candidate Score currently used at West Point was examined.

METHOD

Sample

The incoming USMA Class of 1994 served as the sample for this research. The class was made up of 1338 plebes, of which 1325 participated. Of the 1325, 1164 (88%) were men and 161 were women. The incoming class represented approximately 10% of the total applicants, so that the subjects are a select group, with expected restriction of range on many of the variables.

Questionnaire Development

The complete questionnaire was administered in July, 1990, shortly after their arrival at West Point. Three measures were included in the questionnaire:

Biodata questionnaire. A 73-item biographical data questionnaire was developed for this research. A number of the items or item topics appeared in previous biodata forms (England, 1971; Glennon et al., 1966; Richardson et al., 1985), while others were developed expressly for this research. Items were included if they addressed behaviors or events seen as relating to: (1) the criteria of interest, with leadership performance as the primary criterion, and attrition from USMA as the secondary one; (2) the ABLE temperaments included in the research, especially Dominance; or (3) aspects of military adaptability and other constructs not covered on the version of the ABLE being used. Those falling into the last category included interpersonal style, preference for rugged pastimes, and quality of familial structure and relationships.

There were a number of constraints involved in item development. First, as mentioned above, preference was to be given to objective and verifiable behaviors, even when the conceptual relationship to the constructs was more tenuous. Second, unlike biodata measures used to predict adult success in work situations, the subjects in this research did not have directly applicable "work experience" as either soldiers or commanders, so that the option of fitting items to a detailed job analysis was not feasible. Third, test administration had to be accomplished within tight time constraints, thus forcing the abandonment of numerous potentially useful items.

Hundreds of items were reviewed for potential inclusion in the questionnaire, from which an initial pool of 124 items were developed. Subsequently, 30 items which were perceived as intrusive or likely to generate hostility from the respondents were dropped, which had the effect of minimizing coverage of some temperaments, notably Emotional Stability. The remaining 94-item questionnaire was then shortened to 66 items because of time constraints during a subsequent, cross-validation administration.

An additional seven items came from a 97-item extracurricular activity and sports participation checklist used previously in Air Force research. The checklist asked about leadership roles in 22 different high school extracurricular organizations or activities, and participation and leadership in 25 varsity sports. Because of low variances on a number of activities and sports, as well as cross-validation time constraints, the activities items and 18 of the sports in the extracurricular activities section were dropped from the cross-validation measure and from further analysis. For each of the seven remaining sports items, questions about sport participation, having lettered in the sport, and team captaincy were combined into a single item. Thus, the final 73-item biodata measure was made up of 66 of the potentially best items from the 94-item version, as well as 7 sports items from the 97-item activity and sports inventory.

ABLE. An 88-item version of ABLE was assembled for this research. The measure included the following scales: a 21-item Emotional Stability scale; a 10-item Dependability scale, here composed primarily of items dealing with endorsement of traditional values, as opposed to other forms of ABLE, which also include nondelinquency items in the Dependability construct; a 14-item Work Orientation scale;

a 12-item Dominance scale; and an 18-item Energy scale. An 11-item Validity scale, designed to detect persons whose responses are consistently contaminated with socially desirable and/or dishonest responses (Hough et al., 1990), was also included.

In addition, the selection measures currently used at USMA were included in the research for the purpose of determining the incremental contribution of ABLE and the biodata. The primary measure is the weighted composite called the Whole Candidate Score. However, other WCS components were also evaluated individually against the criteria, in order to isolate the determinants of success on each criterion. These were scores on the SAT (V+M combined), high school rank, the Leadership Potential Score (LPS), and the Physical Aptitude Examination (PAE), all of which were described earlier.

Keying Procedures and Strategies

In keying the biodata, a balance was struck between the rational and empirical approaches. While rational, a priori keying assumes that relationships between item and criteria should be intuitively obvious, an empirical strategy allows for less obvious and more complex relationships to be uncovered. However, an overly empirical approach could lead to the coding of items in illogical ways that are unlikely to be replicated in future samples. For this reason, a number of experienced practitioners commonly use some judgement in empirical keying. Based on consultations with some practitioners, including Mumford (personal communication), the following strategies for logically tempering "dustbowl empiricism" with a more theoretical "rainforest empiricism" (Mael, in press) emerged.

One issue concerns the correct keying of non-continuous items, such as "Which of these courses did you enjoy most?". The experts advised treating each response alternative as a separate item, so that those choosing "Math" were contrasted with all others, as were those choosing "English", "Science", etc. The reasoning is that the exact configuration of the five choices may be too idiosyncratic to be replicated consistently, thus leading to increased shrinkage upon cross-validation. When two or more alternatives form a logical subset, they would of course be combined, and contrasted as a unit to the other options. Thus, a non-continuous item with five alternatives could actually be used as up to five separate items (Hogan & Stokes, 1989).

Another common problem regards items that contain alternatives chosen by few people. For example, in the question "How much sleep do you need per night?", if only 3% of subjects respond "5 hours or less" to the question, the mean associated with that response will likely be unreliable. Therefore, for the present research, alternatives chosen by less than 10% of the sample were considered low frequency alternatives, and treated in one of two ways. If the item was continuous, as in the example above, the low-frequency response was combined with an adjacent response. In this example, the

"5 hours or less" response group would be merged with the "6-7 hours" response group to form one category. With a non-continuous item, low frequency responses were coded "1" and set at the mean, or, in the case of dichotomous coding, set to the same value as the rest of the "other" category. In both these cases, these adjustments would minimize the correlations with the criterion, but would be expected to provide more conservative and stable indications of underlying relationships. Items having overall poor variance (i.e. lacking at least two response choices each endorsed by 10% of the respondents) inevitably did not correlate with any criteria, and therefore had to be dropped completely.

Another issue involves possible illogical keying of items based on strict empiricism. For example, in the item "How many years did you play varsity chess in high school?", suppose that the criterion means for responses on this sample were 2.8 ("not at all"), 3.1 ("1 year"), 3.4 (2 years"), 3.0 (3 years"), and 3.7 ("4 years"). Using a strict empirical key, one would have to assign a lower value to 3-year participation than 1 or 2 year participation. However, barring a compelling post-hoc theory, one would probably assume a sample-specific quirk, especially if the sample was only moderate-sized. Rather than code it this way and incur significant shrinkage, a more logical approach would be to fit this response within the continuum and accept a smaller derivation sample correlation in return for a more stable estimate of true population values.

Keying to ABLE Scales

In the current research, keying to ABLE was empirical, although a good deal of the logical discretion described above was used in assigning weights. Keying of items to each ABLE scale involved several steps. First, means on the ABLE scale for each biodata item response were calculated. Next, a 0, 1, or 2 was assigned to each response alternative. If the response fell within .05 of the mean, it was considered to be at the mean and was assigned a value of 1. Responses with means greater than .05 above the mean were assigned a 2 while responses with means greater than .05 below the mean were assigned a 0. If no responses were more than .05 away from the mean but two heavily-endorsed responses were further than .05 from each other, those responses were coded 0 and 1 or 1 and 2, depending on whether the higher or lower choice was closer to the mean. Based on advice from other practitioners, options were limited to 0, 1, and 2, even if a 4 or 5-point continuum was feasible. Examples of keying items in this manner are presented in Appendix A.

Once all items were coded in this way, they were correlated with each ABLE scale. Items with significant correlations of at least .075 with a scale were used to create each of five ABLE-equivalent biodata scales. This .07-.08 value was indicated by Mumford (personal communication) as generally being the minimum threshold for stability upon cross-validation. The five scales were: Bio-Emotional Stability (22 items); Bio-Dependability (27 items); Bio-Work Orientation (32 items); Bio-Dominance (57

items); and Bio-Energy (40 items). As mentioned above, the item pools for each scale were not mutually exclusive, and no attempt was made to derive factorially distinct scales.

Finally, a biodata composite for the whole ABLE was created. To do this, the best codings of each item, regardless of which ABLE-keyed scale they had come from, were utilized to form a composite, representing the best of the five temperament-keyed scales. The resultant 75 item scale was called Bioabsum.

Criterion measures. Three criterion measures were used for this research. The first was attrition from the initial six-week basic training period known colloquially as "Beast Barracks", which takes place before the onset of classes. The second was ratings of demonstrated leadership capability, which were also collected at the end of the six-week training period. The third criterion was ratings of demonstrated leadership capability, which were collected at the end of the first semester of classes in December, 1990. Although the leadership rating scales for the six-week and fall periods were identical, the moderate correlation between the two measures (r = .35), as well as evidence of differential relationships with the predictors, served as compelling grounds not to combine the ratings or treat them as repeated measures of the same criterion.

RESULTS

ABLE and Bioabsum

Descriptive statistics for the five ABLE scales used in this research are shown in Table 3.

The intercorrelations between the ABLE scales are also shown in Table 3. All ABLE reliabilities were in the acceptable range, and were comparable to those in previous ABLE research.

The correlations between each of the biodata scales keyed to ABLE scales and the ABLE scales appear in Table 4. The correlation between the composite biodata scale Bioabsum and the overall ABLE also appears in Table 4. As can be seen, the correlations between each ABLE scale and its equivalent biodata scale range between .37 and .53. The only off-diagonal correlations between ABLE scales and the biodata scales of other ABLE scales that were of similar magnitude were those between Bio-Dependability and ABLE Work Orientation (r =.50) and between Bio-Energy and ABLE Dominance (r = .39).

Table 3

Descriptive Statistics and Intercorrelations for ABLE scales

Variable	Items	Mean	SD	1	2	3	4	5	6
1. Emotional Stability	21	2.36	.30	.84					
2. Dependability	10	2.54	.28	.18	.70				
3. Work Orientation	14	2.37	.37	.18	.49	.84			
4. Dominance	12	2.53	.32	.36	.17	.33	.82		
5. Energy	18	2.34	.22	.57	.38	.51	.44	.81	
6. ABLE Total	75	2.40	.22	.73	.55	.69	.64	.84	.92

n = 1324. For all correlations, $p < .001$. Alpha coefficient appears in diagonal

Table 4

Intercorrelations Between ABLE-Keyed Biodata Scales and ABLE Scales

Variable	Items	ES	Dep	WO	Dom	EN	ABLE
Bio-Emotional Stability	22	.37	.07*	.17	.27	.33	.37
Bio-Dependability	27	.03#	.42	.50	.15	.22	.33
Bio-Work Orientation	32	.09	.34	.53	.26	.27	.40
Bio-Dominance	57	.20	.11	.27	.49	.29	.38
Bio-Energy	40	.28	.18	.34	.39	.44	.47
Bioabsum	75	.19	.25	.44	.41	.34	.45

n = 1314-1334; # = n.s.; * = $p < .05$; for all others, $p < .001$

It should be noted that the ABLE scales were themselves not orthogonal, with correlations between scales ranging from .17 to .57. In addition, because the same items were used on multiple biodata scales, there were large correlations between some of the biodata scales, with rs ranging between .08 and .85. Thus, some degree of overlap in the off-diagonal coefficients was inevitable. In spite of this, to a great extent the biodata did manage to capture the specific ABLE constructs that they were keyed to, and demonstrated some degree of discrimination in their relationships to the ABLE scales.

Interrelationships Among Predictors

In Table 5, the intercorrelations of the ABLE, Bioabsum, the USMA Whole Candidate Score (WCS), and each of the component USMA predictors (SAT, high school rank, LPS, and PAE) for the cadet sample is shown. It appears that high school rank, perhaps as a correlate of grade-point average, is most strongly related to a cadet's WCS score. However, it must be stressed that these relationships reflect a great deal of range restriction on the variables. In other words, there is no doubt that for the total candidate population, aptitude scores are the greatest determinant of WCS scores, and that the 10% of candidates accepted at USMA had significantly higher SAT scores than the 90% who were rejected. Table 5 reflects relationships within the group that was accepted, and demonstrates that high school rank (and, by inference, GPA) is the greatest determinant of WCS rankings within this select group of acceptees. By the same measure, the ABLE and biodata scores of those who were rejected may also would have been lower and more varied than those of cadets, because these measures were also subject to range restriction.

In this sample, ABLE had a small but significant relationship with all components of the WCS except for SAT. This may be consistent with previous evidence of little overlap between ABLE and ASVAB, the military aptitude test, or it may again be a function of range restriction. Bioabsum showed the largest relationship with LPS, which in part also captures background experiences. Yet, it also relates more strongly to high school rank than either the LPS or ABLE, as would be expected from the wider coverage of topics in the biodata measure. Surprisingly, both Bioabsum and the LPS had negative correlations with SAT scores.

Table 5

Intercorrelations Between ABLE, ABLE-Keyed Biodata (Bioabsum), and Current USMA Predictors

Variable	1	2	3	4	5	6
1. ABLE Total						
2. Bioabsum	.45					
3. WCS	.11	.24				
4. SAT	.01#	-.06*	.21			
5. High School Rank	.09	.22	.71	-.03*		
6. LPS	.13	.37	.12	-.10	.12	
7. PAE	.12	.07*	-.11	-.09	-.11	.09

n = 1314-1334; # = n.s.; * = $p < .05$; ** = $p < .01$; for all others, $p < .001$

Relationships to Six-week Attrition

The correlations between the ABLE scales, the equivalent biodata scales, the USMA predictors, and six-week attrition appear in Table 6. It should be noted that because attrition was a dichotomous criterion measure, and there was less than 10% attrition in the sample, the maximum correlation possible was .55, as explained by Nunnally (1978, p. 146). Nevertheless, each of the ABLE scales was related to attrition, with the relationships for Emotional Stability, Dependability, and Energy Level being highest. Each of the biodata scales was related to attrition as well. None of the ABLE scales had a significantly higher or lower relationship to attrition than the equivalent biodata scale.

Currently, USMA does not utilize any measure to anticipate six-week attrition. Thus, strictly speaking, the WCS was not developed as an indicator of attrition from cadet basic training. However, as it is the current USMA basis for selection, it is of interest to compare the performance of the WCS and its components to that of the ABLE and biodata. In contrast to the ABLE and biodata, the WCS was not related to attrition for this sample. This may be explained by the dominant influence of high school rank, which had a negative, non-significant relationship to attrition, on WCS variance in this sample. The SAT, LPS, and PAE had small, significant relationships with attrition, though they did not equal those of ABLE total or Bioabsum, nor those of the most powerful ABLE and Biodata scales.

Incremental Validity of ABLE and Biodata. A series of multiple regressions was performed to determine the incremental contributions of the ABLE and biodata scales to accounting for attrition over and above the WCS. Both the ABLE and Bioabsum were found to have incremental validity when entered separately. By contrast, the WCS did not account for significant variance when entered with either the ABLE or Bioabsum. When all three were entered together, only the ABLE was individually significant, thus demonstrating considerable overlap between ABLE and biodata when keyed in this fashion.

The contribution of individual ABLE and biodata scales, over and above the WCS, was also evaluated in order to pinpoint the temperaments that played the biggest role in successfully accounting for variance in the criterion. For six-week attrition, each ABLE scale made a significant contribution when entered separately with WCS. However, when entered together, the Emotional Stability and Dependability scales were the only ones to have significant beta weights.

In order to do the same assessment for the biodata scales, it was necessary to first merge scales with extremely high intercorrelations in order to avoid multicollinearity. As noted earlier, the biodata scales shared items and thus inevitably overlapped at times, even though keyed differentially. Thus, Bio-Dependability and Bio-Work Orientation

(r = .85) were merged, as were Bio-Emotional Stability and Bio-Energy (r =.71). When entered separately with WCS, each combined biodata scale, as well as Bio-Dominance, provided incremental value. When entered together, however, only the Bio-Emotional Stability/Energy combination showed a significant individual contribution. When ABLE was also entered, however, no biodata analog was significant.

Relationships to Six-week Leadership Ratings

The correlations of the ABLE scales, the equivalent biodata scales, and the USMA predictors with six-week leader ratings appear in Table 6. Each of the ABLE scales was related to leadership performance, with the relationships for Emotional Stability, Dominance, and Energy Level being highest. In a departure from the other criteria, two of the biodata scales, Bio-Dependability and Bio-Work Orientation, did not have a significant relationship with the criterion. Using the formula for comparing correlations of two variables with a third variable found in Cohen and Cohen (1983, p. 56-57), it was determined that in these two cases, the correlations for the biodata scales were significantly lower than those of the equivalent ABLE scales. Each of the other biodata scales, as well as Bioabsum, was related to attrition, although Bioabsum's relationship with the ratings was clearly pulled down by the inclusion of Bio-Dependability and Bio-Work Orientation items. The result was that Bioabsum's correlation with the criterion was significantly lower than that of ABLE (t_{1184} = 3.21, $p < .01$).

Once again, the WCS was not related to the criterion. The same was true of SAT and high school rank. Conversely, the PAE had a significant relationship to the ratings, while the LPS had a smaller, but still significant, relationship as well.

Incremental Validity of ABLE and Biodata. An identical series of multiple regressions was performed to determine the incremental contributions of the ABLE and biodata scales over and above the WCS. Both the ABLE and Bioabsum provided incremental validity when entered separately. By contrast, the WCS did not account for significant variance when entered with either the ABLE or Bioabsum. When all three were entered together, only the ABLE had a significant value.

Each ABLE scale made a significant contribution when entered separately with WCS. However, when entered together, the Emotional Stability scale was the only one to have a significant beta weight.

Table 6

Correlations of ABLE Scales, Biodata Keyed to ABLE, and USMA Predictors with Attrition and Leadership Criteria

Variable	Six-Week Attrition	Six-Week Leadership	Fall Leadership
Emotional Stability	.12	.16	.02#
Bio-Emotional Stability	.10	.17	.05*
Dependability	.12	.10	.14
Bio-Dependability	.08**	-.01#	.13
Work Orientation	.08**	.10	.14
Bio-Work Orientation	.08**	-.01#	.12
Dominance	.06*	.12	.07**
Bio-Dominance	.09	.07**	.05*
Energy	.11	.18	.06*
Bio-Energy	.10	.15	.08**
ABLE Total	.14	.19	.11
Bioabsum	.11	.07*	.10
WCS	.01#	.04#	.20
SAT	.07**	.01#	.02#
High School Rank	-.04#	-.01#	.17
LPS	.08**	.05*	.07*
PAE	.06*	.17	.02#

n = 1314-1334 (attrition); 1185-1191 (leadership); # = n.s.; * = $p < .05$; ** = $p < .01$; for all others, $p < .001$

When entered separately with WCS, the combined Bio-Emotional Stability/Energy scale and the Bio-Dominance scale each provided incremental value. When entered together, only the Bio-Emotional Stability/Energy combination showed a significant individual contribution. This remained true, however, even when ABLE was also entered. Surprisingly, the biodata Dominance scale had a significant negative relationship to the criterion when entered with WCS, ABLE, and Bio-Emotional Stability/Energy scale, suggesting a possible role as a suppressor.

Relationships to Fall Semester Leadership Ratings

The correlations of the ABLE scales, the equivalent biodata scales, and the USMA predictors with fall semester leader ratings appear in Table 6. Each of the ABLE scales, with the exception of Emotional Stability, was related to fall ratings, with the relationships for Dependability and Work Orientation being highest. The relationships with Dominance and Energy Level were lower than they were with the

other criteria. By contrast, each of the biodata scales, including Bio-Emotional Stability, was related to these leadership ratings. In each case, the relationship to the ratings was comparable to that of the equivalent ABLE scale, with no statistically significant differences. Overall ABLE and Bioabsum also showed their closest proximity to each other with this criterion.

In contrast to the previous criteria, WCS had the strongest relationship of any predictor to the fall ratings. The correlation between high school rank and the fall ratings was also higher than that of any ABLE or biodata scales. The LPS had a small but significant relationship with fall ratings, while the SAT and PAE did not.

Incremental Validity of ABLE and Biodata. The same series of multiple regressions were performed with this criterion. Both the ABLE and Bioabsum were found to add incremental validity when entered separately with WCS. As opposed to the previous criteria, when Bioabsum and ABLE were entered together without WCS, the biodata scale added significantly to the coefficient. Once again, though, when entered with both ABLE and WCS, Bioabsum did not add significant variance. In this instance, WCS also accounted for significant variance when entered with either ABLE or Bioabsum, or with both together. Apparently the redundancy of Bioabsum derives from partial overlap with both ABLE and WCS, rather than from extensive overlap with ABLE.

Three ABLE scales (Dependability, Work Orientation, and Dominance) made significant contributions to WCS when entered separately. However, when entered together, only Dependability had a significant beta weight. When the biodata scales were entered separately with WCS, Bio-Emotional Stability/Energy alone provided incremental value. Entered simultaneously, however, none provided a significant contribution, demonstrating the possibility of still more multicollinearity.

Social Desirability Analyses

Table 7 shows the correlation between the ABLE validity (social desirability detection) scale and each of the ABLE scales, as well as the overall ABLE. The same correlations are shown for the biodata scales keyed to each ABLE scale, and the overall ABLE composite (Bioabsum). In each case, the correlation with social desirability for the ABLE scale was significantly higher than the correlation for the equivalent biodata scale (Emotional Stability, $t_{1324} = 2.30$, $p < .05$; Dependability, $t_{1324} = 2.34$, $p < .05$; Work Orientation, $t_{1324} = 6.12$, $p < .01$; Energy, $t_{1324} = 3.51$, $p < .01$; overall ABLE versus Bioabsum, $t_{1324} = 6.12$, $p < .01$). The sole exception was Dominance, for which both the ABLE and biodata scales had small relationships to the social desirability scale.

Table 7

Correlations of ABLE Scales, Biodata Keyed to ABLE, and USMA Predictors with ABLE Validity (Social Desirability) Scale

Variable	SD Scale	Variable	SD Scale
Emotional Stability	.16	WCS	.04#
Bio-Emotional	.09	SAT	-.03#
Dependability	.31	High School	.09
Bio-Dependability	.25	LPS	.06
Work Orientation	.39	PAE	-.00
Bio-Work	.24		
Dominance	.09		
Bio-Dominance	.07		
Energy	.23		
Bio-Energy	.13		
ABLE Total	.34		
Bioabsum	.18		

n = 1314-1334; # = n.s.; * = $p < .05$; ** = $p < .01$; for all others, $p < .001$

Social desirability was not related to performance on any of the criteria in the research. It was unrelated to six-week attrition ($r = .04$), summer leader ratings ($r = .04$), and fall leader ratings ($r = -.01$). Some previous studies have shown socially desirable responding or faking to be criterion-related. When positively related, it has been interpreted as demonstrating self-esteem (Hogan & Stokes, 1989; Zerbe & Paulhus, 1987), and when negatively related, it has been interpreted as measuring defensiveness and approval-seeking (Crosby, 1990; Crowne & Marlowe, 1960). In either case, it has been treated as meaningful variance by these researchers, rather than measurement error. Clearly, in this sample, socially desirable responding did not account for significant variance in any of the criteria. Furthermore, partialling the effects of those scores from the ABLE and biodata predictor-criterion relationships did not affect those relationships in any way.

Although there was a significant, positive relationship between the social desirability scale and each of the ABLE and biodata scales, there was no relationship between the social desirability scale and either WCS, SAT, or the PAE. However, there was a positive relationship with the LPS. It should be noted that common method variance with the ABLE and biodata scales that appeared in the same instrument with

the validity scale would exaggerate their relationships. Surprisingly, the relationship with high school rank was positive and comparable to that of some of the biodata scales, even though the USMA high school rankings were not derived from self-report sources.

DISCUSSION

The current research effort was conducted under a number of severe constraints. First, because incoming cadets rather than applicants were sampled, the range of variance on all predictor variables was sharply restricted. The attrition criterion measure was also restricted, limiting the maximum correlation coefficient to .55. Second, although typical biodata inventories often have 200-300 items, time constraints in the USMA setting, especially in the cross-validation effort, limited the researchers to 73 items. Third, the item pool was limited to external, objective, and mainly verifiable items in order to render the instrument potentially usable for admissions. Finally, the age and lack of military experience of the cadets made it impossible to measure directly "job-relevant" previous experiences.

In spite of these limitations, the findings of this research were highly encouraging. Five biodata scales were created to parallel temperament scales from the Army ABLE. In each case, the biodata scales showed a clear relationship to the equivalent ABLE scale and almost always a smaller relationship to the other ABLE scales. The biodata scales were also compared with the ABLE scales in their relationship to each of three criterion measures. Out of a total of 15 such comparisons, the biodata measures had a statistically smaller relationship to the criterion in only two cases. In some cases, the biodata scales actually had a slightly higher relationship to the criterion. These results demonstrate that it is possible to develop objective biodata measures that will be substantially analogous to valid temperament measures, even under the aforementioned constraints.

Furthermore, with each criterion, either overall ABLE or Bioabsum (the biodata equivalent) added incremental validity over and above the WCS measure currently used by USMA. For two of the criteria, Bioabsum was redundant with ABLE and did not account for additional variance, while for the third Bioabsum had a lesser overlap. Insofar as the biodata were keyed to maximize their relationship with ABLE scales in this research, this redundancy is desirable. The results do not preclude the possibility that the biodata, keyed directly to the criterion, would show less overlap with ABLE and account for more variance in the criteria. Results of empirical keying were not included in this report because of the need to properly cross-validate empirical keys.

Moreover, another anticipated benefit of using biodata analogs, that of reducing vulnerability to socially desirable responding, was also realized. Four of the five individual biodata scales, as well as the overall biodata scale, had a significantly smaller correlation with the ABLE validity scale than the equivalent ABLE scale. Thus, while socially desirable responding did not seem to contaminate relationships of the

predictors to the criteria, the use of objective biodata does seem to provide a possible minimization of the faking problem. Once again, the present research does not preclude the possibility that biodata keyed directly to the criterion would show less vulnerability to socially desirable responding than either a temperament scale or even the same biodata keyed to a temperament scale. In fact, initial indications from empirical keying of these scales suggest that this is so.

The results are also useful in pinpointing which temperament factors are most important in determining early success at West Point. Cadets who attrited during the six-week training tended to be especially lower in stress tolerance (emotional stability) and in the endorsement of traditional values (dependability). The cadet who succeeds in completing initial training period may tolerate stress, and be willing to accept authority and regimentation, to a greater degree than peers who choose to leave USMA at that stage. Among those who did not attrit, cadets rated highest in leadership performance during the six-week training period were also distinguished most clearly by their greater emotional stability and stress tolerance. Finally, dependability, again in the sense of endorsement of traditional values, had the strongest relationship of all temperaments to leadership behavior in the ratings from the fall semester.

One of the puzzling results of this research was the relatively minor role played by the temperament construct referred to as dominance. Although it was related to both leadership criteria, dominance was never the primary predictor of either training or fall semester leadership ratings. It is possible that because the primary role of the plebe is to be a good team player, rather than to direct other cadets in the accomplishment of their duties, the importance of dominance does not become apparent until later in the cadet's career. It would be important to obtain criterion measures from a cadet's last two years at USMA before dismissing the role of dominance.

The relatively low relationship between summer and fall leadership ratings illustrate that summer and fall performance represent two components of the cadet leader role, rather than repeated measures of the same construct. In support of this premise, it can be seen that the PAE is among the best predictors of summer leadership ratings, yet had a non-significant relationship to fall ratings. Conversely, high school rank is strongly related to fall ratings, yet unrelated to summer ratings. Seemingly, successful summer performance is associated with excellent physical conditioning and stress tolerance, which is also manifested in high energy level, while success in the fall semester relates strongly to previous academic achievement and dependability. However, the degree to which leadership propensity or ability accounts for variance in each of the settings, given the aforementioned lack of actual leadership opportunities for new plebes, is unclear. These discrepant results suggest the need to evaluate the temperament and biodata measures against leadership scores throughout the four-year USMA experience, as well as beyond.

In summary, the results of the current research suggest a useful role for biodata as an indicator of success at USMA. Additional research is needed to substantiate and elaborate on these findings. First, the biodata need to be keyed empirically to each of the criteria, and the results compared to the findings of the current effort. In this way, it can be determined if keying biodata to temperament scales is an optimal or counter-productive use of biodata. Second, the research needs to be replicated, in order to evaluate the stability of the biodata-to-ABLE keys, as well as to cross-validate the empirical keys. This replication is currently underway. Finally, additional effort must be made to relate these temperament and biodata measures to more longitudinal measures of leadership success. In this way, the full value of integrating temperament and biodata into USMA admissions can be determined.

REFERENCES

Arvey, R. D. (1983). Fairness in selecting employees. Reading, MA: Addison-Wesley.

Asher, J. J. (1972). The biographical item: Can it be improved? Personnel Psychology, 25, 251-269.

Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. Personnel Psychology, 27, 519-533.

Barge, B. N. (1987). Characteristics of biodata items and their relationship to validity. Paper presented at the 95th annual meeting of the American Psychological Association, NY, NY.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.

Crosby, M. M. (1990). Social desirability and biodata: predicting sales success. Paper presented at the fifth annual conference of the Society for Industrial and Organizational Psychology, Miami Beach, Florida.

Crowne, D. P., & Marlowe, D. (1960). The approval motive. New York: Wiley.

Dunnette, M. D. (1962). Personnel management. Annual Review of Psychology, 285-313.

England, G. W. (1971). Development and use of weighted application blanks (Rev. Ed.). Minneapolis: University of Minnesota, Industrial Relations Center.

Gandy, J. A., Outerbridge, A. N., Sharf, J. C., and Dye, D. A. (1989). Development and initial validation of the Individual Achievement Record. Washington, DC: U.S. Office of Personnel Management.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Glennon, J. R., Albright, L. E., & Owens, W. A. (1966). A catalog of life history items. Greensboro, NC: Creativity Research Institute of the Richardson Foundation.

Guthrie, E. R. (1944). Personality in terms of associative learning. In J. McV. Hunt (Ed.), Personality and the behavior disorders (Vol. 1). New York: Ronald Press.

Henry, E. R. (1965). Research conference on the use of autobiographical data as psychological predictors. Greensboro, NC: The Richardson Foundation.

Hogan, J. B., & Stokes, G. S. (1989). The influence of socially desirable responding on biographical data of applicant versus incumbent samples: Implications for predictive and concurrent research designs. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Boston, MA.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D. & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75, 581-595.

Landy, F. J., & Trumbo, D. A. (1980). Psychology of work behavior. Homewood, Illinois: Dorsey.

Mael, F. A. (in press). A conceptual rationale for the domain and attributes of biodata items. Personnel Psychology.

Mitchell, T.W. & Klimoski, R.J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. Journal of Applied Psychology, 67, 411-418.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. Applied Psychological Measurement, 11, 1-31.

Mumford, M. D., & Stokes, G. S. (1991). Developmental determinants of individual action: Theory and practice in the application of background data. In M.D. Dunnette (ed.) The handbook of industrial and organizational psychology. (2nd edition, pp. 1-78). Palo Alto, CA: Consulting Psychologists Press.

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.) Handbook of Industrial Psychology. New York: Rand McNally.

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. Journal of Applied Psychology, 65, 569-607.

Pace, L. A., & Schoenfeldt, L. F. (1977). Legal concerns in the use of weighted application blanks. Personnel Psychology, 30, 159-166.

Page, D. P. (1934). Measurement and prediction of leadership. American Journal of Sociology, 41, 31-43.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 46, 598-609.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 35, 1-62.

Richardson, Bellows, Henry & Co. (1985). Supervisory Profile Record. Washington, D.C.: Author.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection? Can validities be made generalizable? Journal of Applied Psychology, 75, 175-184.

Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. Journal of Applied Psychology, 75, 569-580.

Schrader, A., & Osburn, H. G. (1977). Biodata faking: Effects of induced subtlety and position specificity. Personnel Psychology, 30, 395-405.

Stricker, L. J. (1987). Developing a biographical measure to assess leadership potential. Presented at the Annual Meeting of the Military Testing Association, Ottawa, Ontario.

Telenson, P. A., Alexander, R. A., & Barrett, G. V. (1983). Scoring the biographical information blank: A comparison of three weighting techniques. Applied Psychological Measurement, 7, 73-80.

Trent, T., Quenette, M. A., & Pass, J. J. (1989). An old-fashioned biographical inventory. Paper presented at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Van Rijn, P. (1980). Biographical questionnaires and scored application blanks in personnel selection. Washington, D.C.: U.S. Office of Personnel Management, Personnel Research and Development Center.

Watson, T. W. (1989). The Leadership Effectiveness Assessment Profile (LEAP): Prototype development and future plans. Paper presented at the annual meeting of the Military Testing Association, San Antonio, TX.

White, L. A., Nord, R. D., & Mael, F. A. (1990). Setting enlistment standards on the ABLE to reduce attrition. Paper presented at the annual Army Science Conference, Durham, NC.

Zerbe, W.J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception: Academy of Management Review, 12, 250-264.

APPENDIX

EXAMPLES OF EMPIRICAL KEYING EMPLOYED IN USMA RESEARCH

APPENDIX

Examples of Empirical Keying Employed in USMA Research

1. Your ranking in your graduating class was?
   Item Mean: 2.365

	Response Mean	Coded
a. in the top 5%	2.456	2
b. in the top 6-10%	2.429	2
c. in the top 11-25%	2.390	1
d. above the 25%	2.311	1
e. we did not have class rankings	1.744	0

2. In your junior year of high school, how many hours have you spent in an average week participating in sports and exercise?
   Item Mean: 2.532

	Response Mean	Coded
a. 5 or less	1.369	0
b. 6-10	2.012	0
c. 11-20	2.390	1
d. 21-30	2.503	1
e. more than 30	2.679	1

3. Were you a member of the debate team in high school?
   Item Mean: 2.460

	Response Mean	Coded
a. yes	2.620	1
b. no	2.012	