PERSONNEL PSYCHOLOGY
1993, 46

RAINFOREST EMPIRICISM AND QUASI-RATIONALITY: TWO APPROACHES TO OBJECTIVE BIODATA

FRED A. MAEL
U.S. Army Research Institute

AMY C. HIRSCH
NYNEX Corporation

Two approaches to objective biodata, designed to achieve the interpretability and stability of rational approaches, yet minimize socially desirable responding, were explored. The first was a quasi-rational attempt to derive biodata analogs to an existing temperament measure, and then use them as rational scales. The second was a theory-based variant of criterion-referenced keying, termed rainforest empiricism. Both were utilized with two consecutive classes of U.S. Military Academy cadets ($n = 2{,}565$) to predict leadership performance over four rating periods. The biodata analogs to the temperament measure added incremental validity over the Academy's current admissions package and had significantly smaller correlations with a social desirability scale than their equivalent temperament scales. Scales developed with the rainforest approach had higher incremental validities and lower social desirability. Both methods demonstrated satisfactory stability upon cross-validation, and provided complementary interpretability. Advantages to each approach, and the implications for their use, are discussed.

Biodata inventories have been the focus of much selection research and practice, in part because of demonstrably higher validities than most other selection techniques (Reilly & Chao, 1982). However, issues involving the two most common strategies (Mumford & Owens, 1987) for assigning weights to each response alternative within an item, known as keying, have yet to be resolved. With the first strategy, empirical keying, weights are typically assigned to each alternative based on its mean score on the criterion of interest, so that the continuum of values within the item is rearranged to reflect scores on the criterion. While purely empirical keying can lead to optimal correlations with criteria, it is highly sensitive to sample-specific characteristics, so that when the key is cross

Portions of this paper were presented at the Seventh Annual Conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec, May, 1992, and at the Annual Meeting of the American Psychological Association, Washington DC, August, 1992. The contributions of William Burke, Jay Silva, and Leonard White to this research are gratefully acknowledged. The views expresses in this paper are those of the authors and do not necessarily reflect the views of the U.S. Army Research Institute, the U.S. Military Academy, or the Department of the Army.

Correspondence and requests for reprints should be addressed to Fred A. Mael, U.S. Army Research Institute, 5001 Eisenhower Avenue, Alexandria VA 22333.

validated, the regression coefficient is vulnerable to excessive shrinkage. It has also been derided as "dustbowl empiricism" for being atheoretical and failing to advance understanding of antecedents of successful performance (Dunnette, 1962; Pace & Schoenfeldt, 1977).

Some researchers instead champion a rational approach in which item alternatives are assigned a priori values based on their theorized relationships to specific constructs (cf. Mitchell & Klimoski, 1982; Stricker, 1988). Rational biodata advocates also claim that their method will result in less shrinkage, an assertion that has received some support (Clifton, Kilcullen, Reiter-Palmon, & Mumford, 1992; Schoenfeldt, 1989), but has also been challenged (Mitchell & Klimoski, 1982). Because the rational approach typically attempts to measure unitary constructs, items that can be clearly related to a single construct, and then combined into homogeneous scales, are preferred. In turn, this should generally lead to a preference for subjective, temperament-like items which can be focused on one, and only one, construct. By contrast, performance of heterogeneous, objective behaviors often draws on multiple individual characteristics, and responses to items about these behaviors are difficult to assign to the influence of a single construct.

Therefore, a possible drawback is that even if items reflecting complex behaviors are used, item alternatives are still coded in terms of single constructs, even if they are a function of multiple influences. For example, a question about having served as class president may be treated as a measure of dominance, and assigned to a dominance scale, though in may be equally indicative of interpersonal charm, academic achievement, and other constructs. In addition, by deciding item directionality a priori, the possibility that a behavior will prove to be beneficial for some outcomes and counterproductive for others is ignored. In applied settings, where tests are likely to be used for multiple purposes and to predict multiple criteria, this could be problematic.

Researchers have also been concerned about the possibility of socially desirable responding and faking on self-report measures (Crowne & Marlowe, 1960; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Shaffer, Saunders, & Owens, 1986). One proposed method of reducing susceptibility to faking has been to limit biodata to objective or even verifiable items (Atwater, 1980; Cascio, 1975). This is problematic for rational keying advocates, in that objective actions tend to be heterogeneous and therefore difficult to attribute to unitary constructs. Thus, limiting biodata to objective items would eliminate the homogeneous and subjective items preferred in the rational approach. What remains is a dilemma about the types of items to be preferred. Those items that may be most resistant to distortion are also perceived as yielding less understanding and as least stable. Conversely, items that may appear to have the most promise for theoretical advance may be most susceptible to distortion, and hence less useful.

In the present research, two different approaches were undertaken in an attempt to bridge the gap between the empirical and rational perspectives. The first was an attempt to gain the conceptual benefit of measuring specific constructs, yet retain the less fakable properties of objective biodata, by using a quasi-rational approach. Objective biodata items were empirically keyed directly to temperament scales, and then used rationally, as scales with a priori values. The goal was to determine if biodata scales could parallel individual temperament scales in their relationships to criteria. This approach is rational in the sense that items are not keyed directly to the criterion, yet empirical in the sense that item values are generated by their empirical relationship to an external referent, in this case temperament scales. Hence, the term "quasi-rational." In the current effort, temperament scales and their biodata analogs were compared in terms of their relationships with four measures of leadership performance and their vulnerability to faking.

A second approach was a variation of empirical keying, in which theoretical discretion was used in the development and keying of the items, leading to a more conservative, but hopefully more meaningful derivation key. In addition, patterns of item relationships were monitored, and searched for multifaceted profiles of successful performance across the same series of criteria. This approach, termed "rainforest empiricism" by Mael (1991), seeks to improve upon the dry, sterile, and atheoretical technique which has earned the metaphorically vivid sobriquet of "dustbowl empiricism." The idea behind this approach is not new; it was advocated in essence by researchers such as Cureton, Owens, and Otis, at a 1965 biodata conference (Henry, 1965, p. 5). Moreover, this approach is implicitly endorsed in practice by numerous experienced biodata researchers. Practical differences between the rainforest and dustbowl perspectives are discussed below.

The research also reflects an attempt to respond to an applied problem at the U.S. Military Academy at West Point (USMA). Originally, interest was expressed in using the Army's Assessment of Background and Life Experiences (ABLE), a temperament measure which predicts attrition, training completion, and leadership potential (Hough et al., 1990; White, Nord, Mael, & Young, 1993), to predict leadership potential and performance. But, it was decided that ABLE could prove inappropriate for use in an admissions package, because of the nature of the items and concerns about the ease of choosing socially desirable responses. Instead, an attempt was made to substitute objective biodata items, and then utilize them either with the quasi-rational approach, as analogs to ABLE temperament scales, or with the rainforest approach, as empirically keyed, stand-alone biodata. Other recent efforts at utilizing biodata in the prediction of leadership ratings and leadership potential (Fleishman et al., 1991; Mumford, Zaccaro, Harding, Fleishman, & Reiter-Palmon, in press; Russell & Kuhnert, 1992; Russell, Mattson, Devlin, & Atwater, 1990; Stricker, 1988) have generally taken a rational keying strategy, unfettered by limitations on the use of subjective items. By contrast, the current effort was conducted in an environment that mandated that only objective and primarily verifiable items be used.

In summary, two approaches to objective biodata were evaluated in this study. The first is a quasi-rational approach in which initial-stage empirical keying provides biodata analogs to existing temperament scales, which are then used like rational scales. The second approach is stylistically more similar to traditional empirical keying, except that theoretical considerations play a larger role in choice of items and in keying decisions. This research examines the viability of each approach. Because the two approaches were undertaken with the same data set, direct comparisons in terms of validity, incremental contribution, fakability, and interpretability were also possible.

Method

Sample and instruments. The 1,325 incoming cadets of the USMA class of 1994, of which 1,164 (88%) were men, completed questionnaires in July 1990 upon their arrival at USMA. Attrition throughout the period of research accounts for the progressively smaller criterion sample sizes. For purposes of replication and cross-validation, most measures were administered to the 1,240 incoming cadets of the class of 1995. The following measures were included:

Assessment of Background and Life Experiences (ABLE). An 88-item version of the Army temperament measure ABLE was assembled for this research. The measure included the following scales: a 21-item Emotional Stability scale; a 10-item Traditional Values scale, whose items deal with endorsement of respect for rules, authority figures, and discipline; a 14-item Work Orientation scale; a 12-item Dominance scale; and an 18-item Energy scale. An 11-item Social Desirability scale, designed to detect persons whose responses are consistently contaminated with socially desirable responses (Hough et al., 1990), was also included. ABLE scales have demonstrated reliability over numerous administrations (White et al., 1993). Because of time constraints, ABLE could be administered only to the class of 1994.

USMA measures. The admissions measures currently used at USMA were included in the research for the purpose of determining the incremental contribution of the biodata (Burke, 1992). The primary measure used for selection is the weighted composite called the Whole Candidate Score (WCS), 60% of which is based on an applicant's standardized test scores (i.e., SAT, ACT) and graduating rank in high school; 30% on the Leadership Potential Score, derived from the School Official Evaluation (an evaluation form filled out by high school instructors), and the Candidate Activities Record, a checklist of extracurricular activities and varsity sports; and 10% on scores on the Physical Aptitude Examination. The

Whole Candidate Score and each of its primary components were evaluated both collectively and individually against the criteria.

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, Albright, & Owens, 1966; Richardson, Bellows, Henry, & Co., 1985), while others were developed for this research. Items were included if they addressed behaviors or events hypothesized to be related to: (a) the criteria of interest, with leadership performance at various stages as the primary criterion, and attrition from USMA as a secondary one; (b) the ABLE scales included in the research, especially Dominance; or (c) other aspects of military adaptability seen by subject matter experts as potentially relevant to success during and after a cadet's tenure at USMA, such as membership in varsity sports teams and preference for rugged pastimes. The ultimate aim of the research is longitudinal (i.e., prediction of performance as a commissioned officer after graduation from USMA). Thus, it was deemed inappropriate to gear the biodata items to job-analysis based dimensions of short-term success at USMA. Rather, leadership performance was defined as broadly as possible to cover both short- and long-term roles. Moreover, evaluation of the interrelationship of the criterion measures from different periods, and their differential relationships to various predictors, was seen as a critical component.

To minimize faking, items were limited to those that were historical, external, objective, first-person, and mostly verifiable in principle, attributes described in Mael (1991). Recent empirical studies (Becker & Colquitt, 1992; McManus & Masztal, 1993) demonstrate that following these guidelines results in reduced faking of responses. However, by limiting the instrument to such items, other potentially useful items had to be abandoned.

Criterion measures. Four operational ratings of demonstrated leadership capability, using the same 12-dimension rating form, were utilized as criterion measures for this research. The 12 dimensions on the form were duty motivation, military bearing, teamwork, influencing others, consideration, professional ethics, planning and organizing, delegating, supervising, developing subordinates, decision making, and verbal/written communications. A single summary score, based on a weighted combination of ratings by cadet superiors and a tactical officer, was provided to the researchers for each rating period. The first rating period was the initial 6-week cadet basic training period known as "Beast Barracks," which occurs prior to the onset of classes. The second and third periods were the fall and spring semesters of the first academic year. The fourth rating period was cadet field training, which takes place during the second summer of attendance at USMA. Reliability and construct validity information on the USMA criteria were unavailable. The moderate correlations between the criterion measures (ranging from .23

to .40), as well as evidence of differential relationships with the predictors, served as compelling grounds not to initially combine the ratings.

Keying Procedures and Strategies

In keying the biodata, a balance was struck between the rational and empirical approaches. While a priori keying was found to be inappropriate with heterogeneous items, it was felt that a strictly empirical approach could lead to coding items in illogical ways unlikely to be replicated with future samples. For this reason, most experienced practitioners commonly use some judgment in empirical keying. Based on consultations with a number of these practitioners, the following strategies for logically tempering "dustbowl empiricism" with a more theoretical "rainforest empiricism" (Mael, 1991) emerged.

One 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 strict empirical keying, one would assign a lower value to 3-year participation than 2- or 4-year participation. However, barring a compelling post-hoc theory, one would be suspicious of a sample-specific quirk, especially if the sample size was only moderate. Rather than code it this way and incur significant shrinkage, a more logical approach would be to fit this response within the continuum, perhaps by assigning a single value to 1–3 years participation. With this strategy, the researcher accepts a smaller correlation within the derivation sample, in return for a more stable and better estimate of true population values.

Another issue concerns the keying of a non-continuous item, 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," and so on. 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 could be combined, and contrasted as a unit to the other options. Thus, this sample item could actually serve as multiple items (Hogan & Stokes, 1989).

Another problem regards items with alternatives chosen by few people. For example, in the question "How much sleep do you need per night?" if only 3% respond "5 hours or less" to the question, the mean associated with that response will likely be unreliable. Thus, following the experts' advice, 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, the low-frequency response was combined with an adjacent response. Thus, in the above 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 set at the mean. While these adjustments would minimize initial correlations with the criterion, it was hoped they would 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 respondents) inevitably did not contribute to validity, and were dropped. In summary, the two ways in which the rainforest approach departs from dustbowl empiricism in this study involve the choice of items and keying of items, at each step attempting to follow Pace and Schoenfeldt's advice to "insure that the 'hand of reason' has been injected into the process" (1977, p. 165).

Biodata analogs to ABLE. In the current research, keying to ABLE was empirical, though tempered by the logical discretion described above. Each biodata item was keyed to each ABLE scale, and, if significantly correlated with the scale, was retained as part of a biodata analog of that construct . Because of dummy-coding of categorical items, the total number of potential items was greater than 73. The five developed scales were: Bio-Emotional Stability (22 items); Bio-Traditional Values (27 items); Bio-Work Orientation (32 items); Bio-Dominance (57 items); and Bio-Energy (40 items). Because of the heterogeneous nature of the objective biodata items, the item pools for each scale were not mutually exclusive, and no attempt was made to derive mutually exclusive scales. Examples of some of the behaviors making up each of these scales are shown in Figure 1.

Rainforest empirical keys. With this approach, biodata scales were created which were empirically keyed to each of the four criteria, using the same rationally-bounded methodology described above. The result was four predictor measures: Bio-Basic (22 items), Bio-Fall (28), Bio-Spring (31), and Bio-Field (26). Each measure was cross-validated through triple cross-validation (Silverhart, 1990). First, each sample was divided into thirds. A series of keys was developed on each possible combination of two-thirds and then cross-validated on the hold-out third. After this round robin was completed, the average of the three derivation validities and cross-validities for each scale was computed. In addition to within-class cross-validation, rainforest keys developed on the whole class of 1994 sample were cross-validated against class of 1995 criterion ratings. Similarly, keys developed on the class of 1995 were utilized with class of 1994 data and criteria, for the purpose of cross-validation. The two methods of cross-validation have complementary appeals. The across-class method minimizes issues of within-sample sampling error (Murphy, 1984), and can indicate stability even in the presence of substantive sample differences. In this case, the classes entered USMA before and after the Desert Storm war, respectively, which could have impacted the decisions of applicants to enter the academy. Within-class

Cadets high in Emotional Stability:
- watch less television
- when angry, exercise or walk vs. fighting or stewing

Cadets high in Traditional Values:
- attend prayers more often
- spend more nights at home each week

Cadets high in Work Orientation:
- highest high school grades and class rank
- ran for elective office more often

Cadets high in Dominance:
- most likely to have been varsity sport team member or captain
- most likely to have held supervisory work position

Cadets high in Energy Level:
- work more hours at part-time jobs
- need less sleep and sleep less on an average night

Figure 1. **Sample Behaviors from the Five Quasi-Rational Biodata Scales**

cross-validation can be done simultaneously, and also negates spurious sample differences in testing time and conditions, as noted below. For this reason, both methods were used.

Results

Comparison of ABLE and Biodata Analogs

Correlations among the ABLE scales and the biodata analogs to the ABLE scales appear in Table 1. The correlations between each ABLE scale and its equivalent biodata scale range between .37 and .53. Only two off-diagonal correlations between ABLE scales and the biodata analogs of other ABLE scales were of similar magnitude.

Note that the ABLE scales were themselves not orthogonal, with correlations between scales as high as .57. And, because the same items were used in multiple biodata scales, some overlap in the off-diagonal coefficients was inevitable. Yet, to a great extent the biodata scales approximated the specific ABLE constructs they were keyed to, and demonstrated discrimination in their relationships to the other ABLE scales.

Relationships to Leadership Ratings

Validation of the biodata analogs. The correlations of the ABLE scales, their biodata analogs, and the USMA predictors, with each of the four criterion measures, appear in Table 2. Each of the ABLE scales was related to leadership performance during basic training and field

TABLE 1

Intercorrelations of Able Scales and Biodata Analogs, Class of 1994

Variable	Items	1	2	3	4	5	6	7	8	9	10
1. Emotional stability	21	.84									
2. Traditional values	10	.18	.70								
3. Work orientation	14	.18	.49	.84							
4. Dominance	12	.36	.17	.33	.82						
5. Energy	18	.57	.38	.51	.44	.81					
6. Bio-emotional stability	22	.37	.07	.17	.31	.34	.54				
7. Bio-traditional values	27	.03	.42	.50	.16	.22	.08	.65			
8. Bio-work orientation	32	.09	.34	.53	.29	.27	.22	.85	.70		
9. Bio-dominance	57	.20	.11	.27	.52	.29	.48	.30	.52	.71	
10. Bio-energy	40	.28	.18	.34	.43	.44	.71	.35	.53	.66	.56

Note: Correlations above .058 are significant at $p < .01$. Alpha coefficients appear in diagonal. Correlations between ABLE scales and their biodata analogs are underlined.
$n = 1{,}324$

TABLE 2

Correlations of ABLE Scales, Biodata Keyed to ABLE, and USMA Predictors with Four Leadership Ratings, Class of 1994

Variable	Basic	Fall	Spring	Field
Emotional stability	.17*	.05	.05	.15*
Bio-emotional stability	.17*	.06	.09*	.23*
Traditional values	.11*	.17*	.16*	.09*
Bio-traditional values	.02	.24*	.27*	.08*
Work orientation	.11*	.19*	.28*	.12*
Bio-work orientation	.02	.23*	.27*	.10*
Dominance	.12*	.07*	.02	.17*
Bio-dominance	.08*	.09*	.09*	.21*
Energy	.18*	.07*	.11*	.20*
Bio-energy	.16*	.09*	.11*	.20*
Whole candidate score	.07*	.32*	.35*	.10*
SAT	.02	.09*	.10*	−.04
High school rank	.03	.30*	.30*	.04
Leader potential score	.07*	.11*	.16*	.16*
Physical aptitude exam	.16*	−.02	.03	.18*

Note: Underlined coefficients are significantly higher than their ABLE or biodata equivalents.
$n = 1{,}183$ (Basic training and Fall); 1,145 (Spring); 1,076 (Field training)
*$p < .05$

training. However, Emotional Stability was not related to fall or spring ratings, nor was Dominance related to spring ratings.

Among the biodata analogs, Bio-Dominance and Bio-Energy were related to each of the criteria, while the other analog scales were related

TABLE 3

Rainforest Biodata Validities and Cross-Validities for Leadership Ratings, USMA Classes of 1994 and 1995

Empirically keyed biodata	Leadership ratings			
	Basic	Fall	Spring	Field
Triple cross-validation within class of 1994	.38 (.18)	.40 (.30)	.42 (.29)	.36 (.28)
	.32 (.21)	.41 (.30)	.45 (.30)	.36 (.30)
	.31 (.29)	.44 (.28)	.45 (.32)	.39 (.21)
Average derivation and cross-validity using triple cross-validation, class of 1994	.34 (.23)	.41 (.30)	.44 (.30)	.37 (.27)
Validity and across-class cross-validation based on keying to total 1994 sample	.30 (.17)	.39 (.17)	.40 (.32)	.34 (.14)
Validity and across-class cross-validation based on keying to total 1995 sample	.31 (.19)	.34 (.22)	.40 (.29)	.28 (.23)

Note: All values in parentheses are cross-validities.
*$p<.01$

to three of the four criteria. In only 2 of 20 comparisons was the coefficient for an ABLE scale significantly higher than its biodata analog, as computed by the formula for testing differences between dependent correlations described in Cohen and Cohen (1983, p. 56–57). Conversely, the biodata analog was significantly higher in 4 cases.

The relatively minor coefficients for Dominance and Bio-Dominance, despite their conceptual relationship to leadership, was explained by USMA researchers. The primary role of the plebe (freshman) during basic training is to be a good team player, rather than to direct other cadets, while fall and spring semester ratings are given in the context of academic accomplishments. Thus, the importance of dominance only emerges at the first real opportunity for leading, field training.

Rainforest validities and cross-validities. Table 3 shows the validities and cross-validities of the rainforest biodata keys for each of the four leadership criteria. For each criterion, the results of the triple cross-validation performed within the class of 1994 are shown, followed by the average derivation and cross-validation score. Then, the validity for the biodata keyed to the total class of 1994 sample, together with cross-validities in parentheses, is shown. Cross-validities were obtained by using the class of 1994 keys with the data from the class of 1995 sample. The final row shows the reverse (i.e., the validity of the keys developed on the 1995 sample and their cross-validities with the data from the class of 1994 sample).

The results in Table 3 show that the cross-validities of the rainforest-keyed biodata were significantly related to the criteria, and did not show excessive shrinkage. In comparing the two types of cross-validation, the within-class method presented a picture of greater stability than did the more conservative, across-class method.

TABLE 4

Correlations of Biodata Keyed to ABLE and USMA Predictors with Four Leadership Ratings, Class of 1995

Variable	Basic	Fall	Spring	Field
Bio-emotional stability	.15*	.04	.03	.11*
Bio-traditional values	.07*	.23*	.29*	.11*
Bio-work orientation	.07*	.24*	.28*	.12*
Bio-dominance	.14*	.09*	.07*	.16*
Bio-energy	.14*	.04	.09*	.12*
Whole candidate score	.17*	.29*	.32*	.15*
SAT	.07*	.07*	.10*	.02
High school rank	.11*	.28*	.30*	.10*
Leader potential score	.18*	.25*	.22*	.17*
Physical aptitude exam	.21*	.10*	.09*	.16*

n = 1,192 (Basic training); 1,162 (Fall); 1,118 (Spring); 1,092 (Field training)
*$p < .05$

Two points need to be considered when assessing the degree of shrinkage in across-class cross-validities. First, the 1995 sample had sufficient missing data on 8 of the 73 items to require dropping those items. In addition, variation between the two classes due to sampling can be evidenced from differences in the correlations between the USMA predictors and the criteria for each of the classes. To the extent that substantive differences between the classes exist, shrinkage cannot be blamed solely on biodata item instability.

Cross-validation of the biodata analogs. ABLE was not administered to the class of 1995. However, biodata analogs, created by keying items to ABLE scales on the class of 1994 data, were utilized with the class of 1995 in order to cross-validate that approach as well. The validities of the biodata analogs and the USMA predictors for the class of 1995 appear in Table 4.

Validities for the class of 1995 were virtually identical to those of the first sample, with only one coefficient significantly lower than its equivalent for the class of 1994. This indicates that shrinkage was not a problem with this method.

Incremental validity of the biodata. In order to determine the relative merits of the two approaches in terms of incremental contribution to prediction with the Whole Candidate Score, a series of hierarchical multiple regressions were performed. These appear in Tables 5 and 6 for the classes of 1994 and 1995, respectively. For the quasi-rational approach, the unit-weighted sum of the five biodata analogs was entered after the USMA Whole Candidate Score, and is referred to collectively as BioABLE. For the rainforest approach, the across-class cross-validities, representing the more conservative assessment of validity, were entered after the Whole Candidate Score.

TABLE 5

Comparison of the Quasi-Rational (BioABLE) and Rainforest Empiricism Approaches in the Prediction of Four Leadership Criteria, USMA Class of 1994[1]

Quasi-rational (BioABLE)				Rainforest empiricism			
	β	*MR*	R^2		β	*MR*	R^2
Basic training							
Whole candidate score	.05	.07*		Whole candidate score	.06*	.07*	
BioABLE	.09*	.11*	.01*	Bio-basic	.19*	.20*	.03*
Fall semester							
Whole candidate score	.29*	.32*		Whole candidate score	.28*	.32*	
BioABLE	.12*	.34*	.01*	Bio-fall	.13*	.34*	.02*
Spring semester							
Whole candidate score	.31*	.35*		Whole candidate score	.28*	.35*	
BioABLE	.17*	.39*	.03*	Bio-spring	.19*	.39*	.03*
Field training							
Whole candidate score	.05*	.10*		Whole candidate score	.08*	.10	
BioABLE	.20*	.22*	.04*	Bio-field	.22*	.24*	.05*

Note: β indicates standardized beta values; *MR* refers to multiple correlations; R^2 is change in R^2.

[1] Rainforest values are from across-class (1995) cross-validities. Multiple *R* values for triple cross-validation method are .24 (Basic training), .38 (Fall), .40 (Spring), and .28 (Field training).

*$p<.05$

TABLE 6

Comparison of the Quasi-Rational (BioABLE) and Rainforest Empiricism Approaches in the Prediction of Four Leadership Criteria, USMA Class of 1995[1]

Quasi-rational (BioABLE)				Rainforest empiricism			
	β	*MR*	R^2		β	*MR*	R^2
Basic training							
Whole candidate score	.14*	.17*		Whole candidate score	.11*	.17*	
BioABLE	.11*	.90*	.01*	Bio-basic	.18*	.23*	.03*
Fall semester							
Whole candidate score	.26*	.29*		Whole candidate score	.23*	.29*	
BioABLE	.11*	.31*	.01*	Bio-fall	.20*	.35*	.03*
Spring semester							
Whole candidate score	.28*	.32*		Whole candidate score	.24*	.32*	
BioABLE	.14*	.35*	.02*	Bio-spring	.24*	.39*	.05*
Field training							
Whole candidate score	.12*	.15*		Whole candidate score	.14*	.15	
BioABLE	.13*	.19*	.02*	Bio-field	.13*	.20*	.02*

Note: β indicates standardized beta values; *MR* refers to multiple correlations; R^2 is change in R^2.

[1] Quasi-rational keys were developed on 1994 data. Rainforest values are from across-class (1994) cross-validities.

*$p<.05$

In each case, the biodata added incremental validity to the Whole Candidate Score, regardless of keying method. A greater contribution of the rainforest approach was evidenced in the prediction of basic training ratings, and to a lesser extent in prediction of field training ratings. If the within-class triple cross-validation method were used instead, the rainforest scales would demonstrate a more pronounced improvement in incremental validity (footnote, Table 5). There were no differences in prediction of fall and spring semester for the class of 1994. The results in Table 6, which utilize cross-validities of both quasi-rational measures and rainforest measures that were developed on the class of 1994 and applied to the class of 1995, may be a fairer comparison. In this case, the rainforest approach had somewhat higher incremental validities in all but field training ratings for class of 1995.

Though not shown in Table 5, additional regressions were performed for the class of 1994 which included the five ABLE scales as well as the biodata and Whole Candidate Score. This was done to test the hypothesis that the ABLE and biodata analog scales would share more overlapping criterion variance than would the ABLE scales and rainforest keys. This was confirmed. The contribution of the biodata analogs to incremental validity over and above that of the Whole Candidate Score and the ABLE scales was non-significant for two criteria (basic training and spring semester), marginal ($p < .054$) for fall semester ratings, and significant only for field ratings. The rainforest keys, however, provided significant incremental validity for each criterion. In each case, the multiple correlation was higher when using the rainforest keys rather than the biodata analogs.

Dimensionality of the Leadership Criterion

The intercorrelations between the four criteria, and the intercorrelations between the four rainforest keys, for both the classes of 1994 and 1995, appear in Table 7. The correlations between the biodata keys in the two field settings, Bio-Basic and Bio-Field, for each class, were .72 and .70, respectively. Similarly, the correlations between the keys for the two academic year ratings, Bio-Fall and Bio-Spring, were .79 and .68, respectively. By contrast, all other correlations among the empirical keys were no higher than .30 for either of the classes. The implication is that there are two dimensions of leadership as rated at USMA, each represented by a distinguishable constellation of items. Although the ratings themselves were unifactorial, and may represent a degree of halo effect, the behaviors and experiences that *predict* each dimension are demonstrably different.

Further analysis of the relationships between the USMA predictors and the rainforest scales provide some support for this premise. High school rank, a measure of past scholastic performance, was related to the Bio-Fall and Bio-Spring scales, but was negatively related to the two field predictors, Bio-Basic and Bio-Field. This pattern was repeated with the Whole Candidate Score, which was dominated by previous academic

TABLE 7

Correlations Among the Four Leadership Criteria for Classes of 1994 and 1995

Variable	Basic	Fall	Spring
Basic training			
Fall semester	.38 (.43)		
Spring semester	.28 (.32)	.41 (.41)	
Field training	.41 (.38)	.23 (.28)	.31 (.28)

Correlations Among the Four Rainforest Keys for Classes of 1994 and 1995

Variable	Bio-basic	Bio-fall	Bio-spring
Bio-basic			
Bio-fall	.22 (.06)		
Bio-spring	.28 (.11)	.79 (.68)	
Bio-field	.72 (.70)	.27 (.19)	.24 (.31)

Note: Coefficients for class of 1995 are in parentheses; all correlations are significant at $p<.05$.

achievement. Conversely, the Physical Aptitude Exam was related to the biodata predictors of field success, but not to the predictors of academic year leadership. Similar support may be derived from correlations involving the ABLE scales. Three constructs that would appear to be most important for interpersonal leadership (Emotional Stability, Dominance, and Energy Level) were related to the Basic and Field ratings, and to their rainforest predictors, to a greater extent than to the academic year ratings. Conversely, Traditional Values and Work Orientation, constructs that relate more to following instructions, steadfastness, and work habits, were somewhat more related to the academic year ratings and their rainforest predictors.

In reviewing the items making up these rainforest scales (utilizing the scales from both classes), patterns emerge for each dimension. For the field-related biodata scales, the dominant themes are indicators of athleticism and fitness, rugged and self-sufficient pastimes, and work experiences during high school including summer, volunteer, and supervisory work. For the academic-year ratings, the main behavioral themes included a range of school-related leadership positions, academic accomplishments, and more time devoted to homework, as well as similar work experience. Thus, both dimensions appear to include situationally appropriate elements of leadership potential, albeit different ones, as well as more general skills and experiences.

Social Desirability Analyses

Table 8 shows the correlation between the ABLE Social Desirability Detection scale and each ABLE and Biodata scale, as well as the USMA predictors. The Social Desirability scale's correlation was significantly higher with each ABLE scale than with its biodata analog [Emotional

TABLE 8

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

Variable	SD scale	Variable	SD scale
Emotional stability	.16*	Whole candidate score	.04
Bio-emotional stability	.09*	SAT	−.03
Traditional values	.31*	High school rank	.09*
Bio-traditional values	.25*	Leader potential score	.06*
Work orientation	.39*	Physical aptitude exam	−.00
Bio-work orientation	.24*	Biodata keyed to:	
Dominance	.09*	CBT	.01
Bio-dominance	.07*	Fall	.11*
Energy	.23*	Spring	.13*
Bio-energy	.13*	CFT	.04

*$p<.05$; $n = 1{,}325$

Stability ($t_{1324} = 2.30$, $p < .05$); (Traditional Values, $t_{1324} = 2.34$, $p < .05$); (Work Orientation, $t_{1324} = 6.12$, $p < .01$;) (Energy, $t_{1324} = 3.51$, $p < .01$.)], with the sole exception of Dominance.

Correlations between two of the measures, Bio-Basic and Bio-Field, and Social Desirability, were even smaller than those of the ABLE analogs, and were not statistically significant. The correlations of Bio-Fall and Bio-Spring with Social Desirability were comparable to the lowest of the biodata analogs, and significantly lower than those of the ABLE scales.

High school rank, which covaried with Bio-Fall and Bio-Spring, had a comparable relationship with the Social Desirability scale, even though it was not derived from self-report. By contrast, the Physical Aptitude Exam, which paralleled Bio-Basic and Bio-Field in pattern of prediction, had no relationship to Social Desirability.

Discussion

In this paper, two different approaches were used in an attempt to achieve the benefits associated with biodata, while minimizing some of the potential pitfalls associated with both the pure empirical and rational perspectives. In the quasi-rational approach, a set of biodata items were keyed to five temperament scales, in order to form a biodata analog to each temperament scale. A second approach was a modification of criterion-referenced keying, termed "rainforest empiricism," in which item choice and empirical keying were tempered with the "hand of reason." This research was conducted under a number of constraints. First, the sample was a highly select group of incoming cadets, who were chosen based on numerous additional considerations besides the explicit predictor. Thus, there may have been additional, unique restriction of range on the biodata. Second, although many biodata inventories often begin pilot testing with 200–300 items, the researchers were constrained to 73 items, possibly hampering coverage of the desired domains. Third, the longitudinal scope of the research, focused ultimately on all stages of a cadet's career, precluded using job analysis of either short-term or long-term roles as the main basis for item generation. Finally, the youth and inexperience of the cadets limited the number of objective, directly "job relevant" previous experiences available for the item pool.

Nonetheless, these findings are encouraging. With the quasi-rational approach, the biodata analogs were compared to the ABLE scales in relation to each of four criteria. Of 20 such comparisons, the biodata measures had a smaller relationship to the criterion in only 2 cases, and in 4 cases had higher relationships. Results with a second, cross-validation sample were comparable. Similarly, with the rainforest approach, the validities and cross-validities of each empirical key were significantly related to the criteria, without evidencing excessive shrinkage.

Further, for each criterion, the biodata analogs added incremental validity over and above the Whole Candidate Score, USMA's current admissions composite. The rainforest keys made greater contributions to incremental validity for the field leadership criteria, and, depending on the sample and cross-validation method, equivalent or greater contributions for the scholastic year criteria. Another difference between the approaches involved the relative contribution of the biodata when both the Whole Candidate Score and ABLE were included in the regression. The biodata analogs developed to approximate ABLE scales were primarily redundant with them, and accounted for additional variance in only one of the four criteria. By contrast, the rainforest scales provided incremental validity in all cases.

Insofar as the biodata analogs were keyed to maximize their relationship to ABLE, their redundancy with the ABLE scales was both expected and desirable. These results demonstrate that it is possible to develop biodata measures consisting of objective items that could parallel more subjective construct measures, should use of subjective measures be unfeasible. In addition, the quasi-rational approach can inform about the behavioral correlates of dispositional constructs, a role previously envisioned for biodata (Henry, 1965, p. 14). Conversely, an implication of the higher incremental validities with the rainforest keys is that although these items could be utilized to measure a priori temperament constructs, some unique variance attributable to the complex, multi-dimensional components of these behaviors could be lost by using that approach. In situations where both the biodata and temperament measures could be used in a selection battery, clearer distinction between the two types of predictors could yield the best prediction.

Another anticipated benefit of using objective items, reducing vulnerability to socially desirable responding, was also realized. Four of the five individual biodata analogs had significantly smaller correlations with social desirability than their ABLE counterparts, and the rainforest-keyed scales had even lower correlations. This appears to

1C

corroborate findings that historical, objective, and verifiable items are less amenable to faking (Becker & Colquitt, 1992; McManus & Masztal, 1993). Another explanation might be that while the ABLE scales were item-keyed, with all options scored on an a priori continuum, the biodata were option-keyed, an approach shown to have lower social desirability (Kluger, Reilly, & Russell, 1991). However, given that objective items inherently lend themselves more to option-keying, the explanations are probably not independent.

Interpretability of Empirically-Keyed Biodata

The most common complaint against empirical keying, non-interpretability, seems inappropriate not only to the quasi-rational approach, but also to the rainforest approach. Multiple keyings of the same items against similar criteria allowed the emergence of topically informative item clusters, and provided a basis for understanding the antecedents of success in both components of early USMA leadership ratings. The relationship between the type of behaviors that were consistently predictive and the criteria can be explained from the perspectives of both the ecology model (Mumford & Owens, 1987; Mumford & Stokes, 1991) and social identity theory (Mael, 1991; Turner, 1987). For example, persons involved in rugged outdoor activities would generally find those endeavors rewarding and reinforcing, and would be more likely than others to excel in military field roles requiring similar activities. Also, a person involved in those activities would be more likely to identify with others sharing similar affinities, an identification that would express itself in comportment and other prototypical components of the rugged lifestyle. Conversely, those receiving high academic semester leadership ratings demonstrated a greater predilection for academic achievement. It is notable that while this group also had a history of leadership roles, it expressed itself primarily in structured, elected roles within the school, rather than in roles signifying emergent leadership. Certainly, a small minority of biodata items that were predictive, such as birth order or size of one's hometown, are not easily included in unitary construct scales or behavioral clusters. Moreover, grouping these items as "background experiences" or "childhood demographics," as is sometimes done, is conceptually meaningless. On the other hand, items like these often have extensive theoretical and empirical literatures devoted to their significance, so that they have the capacity, even in isolation, to be informative about the qualities required of those who are successful on a given criterion. A single, reliable measure of an objective experience or behavior can be a critical indicator of a whole syndrome of behaviors leading to, or flowing from, that single event.

Comparing the Approaches

There are practical considerations involved in the use of either of these approaches. Because heterogeneous items and empirical keying are utilized at the initial stage of each approach, larger item and sample pools than might be necessary with a pure rational approach would be desirable. For the quasi-rational approach, it is necessary to specify the temperament or attitude predictors to be linked with biodata analogs prior to item generation. Because the temperament items must also be administered, they actually compete with potential biodata items for inclusion in an instrument administered under time constraints. Thus, it would be prudent to locate the most concise, orthogonal, and reliable measures for desired constructs. Were one to attempt to develop mutually exclusive biodata analogs, unlike the current effort, an even larger initial item pool would be needed, in the hope of finding objective, yet homogeneous, items. If the testing environment does not require objective items (Mael, 1991), it should be easier to develop mutually exclusive scales. White and Kilcullen (1992) have in fact developed biodata analogs to ABLE with a rational approach, albeit with inclusion of subjective items, that also demonstrated similar validities to their ABLE counterparts, as well as lower Social Desirability.

The quasi-rational approach is less labor intensive and easier to use in the short term. After initial keying, the scales can be used much like a priori, rational scales, without further keying. However, as Mael (1991) has argued, extended application of the rainforest model should also result in items and item-clusters with known properties and nomological networks, which optimally could also be used without further keying.

To a certain extent, preference for each method may be a matter of taste. For the researcher who is most comfortable with preconceived psychological constructs, the quasi-rational approach assures the type of interpretability to be expected before embarking on testing. Conversely, the clusters and constructs yielded by the rainforest approach often stray from trait and dispositional concepts, and can straddle other domains of psychology, as well as sociology and related social sciences. A similar digression from traditional categories led Fleishman (1988) to call the subgrouping methodology of Mumford and Owens (1987), which is analogous to lifestyle segmentation in market research (Mitchell, 1983; Plummer, 1974), a novel and important trend in personnel selection research. Thus, it could be argued that the quasi-rational approach, much like the pure rational approach, achieves greater theoretical parsimony. In return, it sacrifices some of the richness of information embedded in the behaviors and experiences measured by high-quality objective items. It would be possible, however, to utilize the same items in both ways, and thereby gain understanding from multiple perspectives.

In summary, the results suggest two potential methods of utilizing biodata to predict and understand behavior, while still retaining the less fakable properties associated with objective biodata. In the current research, information from both approaches provided complementary insight into the types of cadets who were most successful in their early leadership roles. Both approaches could yield even better validities under more optimal selection conditions. These would include more specific jobs, which would allow for detailed job analysis and better defined criteria, and longer testing time, allowing for a larger initial item pool. A larger item pool would also make it possible to compare the quasi-rational and rainforest forms with pure rational and pure dustbowl techniques. In addition to following the cadet samples through additional criteria, additional research with these approaches, such as widening the scope of constructs utilized and the types of populations sampled, is being conducted. Both approaches should be given consideration for their potential role in upgrading the stature and utility of objective biodata.

REFERENCES

Atwater DC. (1980, May). *Faking of an empirically keyed biodata questionnaire*. Paper presented at the Annual Meeting of the Western Psychological Assoc. Convention.

Becker T, Colquitt A. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. PERSONNEL PSYCHOLOGY, *45*, 389–406.

Burke WP. (1992). *Admissions variables through the class of 1995*. Office of Institutional Research (Research Report 92-003). West Point, NY: U.S. Military Academy.

Cascio WF. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, *60*, 767–769.

Clifton TC, Kilcullen RN, Reiter-Palmon R, Mumford MD. (1992, August). *Comparing different background data scaling procedures using triple cross-validation*. Paper presented at the Annual Convention of the American Psychological Association, Washington, DC.

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

Crowne DP, Marlowe D. (1960). *The approval motive*. New York: Wiley.

Dunnette MD. (1962). Personnel management. *Annual Review of Psychology*, *13*, 285–313.

England GW. (1971). *Development and use of weighted application blanks* (rev. ed.). Minneapolis: University of Minnesota, Industrial Relations Center.

Fleishman EA. (1988). Some new frontiers in personnel selection research. PERSONNEL PSYCHOLOGY, *41*, 679–702.

Fleishman EA, Mumford MD, Zaccaro SJ, Levin MY, Korotkin A, Hein MB. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and cognitive interpretation. *Leadership Quarterly*, *2*, 245–287.

Glennon JR, Albright LE, Owens WA. (1966). *A catalog of life history items*. Greensboro, NC: Creativity Research Institute of the Richardson Foundation.

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

Hogan JB, Stokes GS. (1989, April). *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 Fourth Annual Conference of the Society for Industrial and Organizational Psychology, Boston, MA.

Hough LM, Eaton NK, Dunnette MD, Kamp JD, McCloy RA. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 581–595.

Kluger AN, Reilly RR, Russell CJ. (1991). Faking biodata tests: Are option-keyed tests more resistant? *Journal of Applied Psychology, 76*, 889–896.

Mael FA. (1991). A conceptual rationale for the domain and attributes of biodata items. PERSONNEL PSYCHOLOGY, *44*, 763–792.

McManus MA, Masztal JJ. (1993, April). *Attributes of biodata: Relationships to validity and socially desirable responding*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Mitchell A. (1983). *Nine American lifestyles: Who we are and where we are going*. New York: Macmillan.

Mitchell TW, Klimoski RJ. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology, 67*, 411–418.

Mumford MD, Owens WA. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11*, 1–31.

Mumford MD, Stokes GS. (1991). Developmental determinants of individual action: Theory and practice in the application of background data. In Dunnette MD, Hough LM (Eds.), *The handbook of industrial and organizational psychology* (2nd ed., pp. 1–78). Orlando, FL: Consulting Psychologists Press.

Mumford MD, Zaccaro SJ, Harding FD, Fleishman EA, Reiter-Palmon R. (in press). *Cognitive and temperament predictors of executive ability: Principles for developing leader capacity* (Final Report). Alexandria, VA: U.S. Army Research Institute.

Murphy KR. (1984). Cost-benefit considerations in choosing among cross-validation methods. PERSONNEL PSYCHOLOGY, *37*, 15–22.

Pace LA, Schoenfeldt LF. (1977). Legal concerns in the use of weighted application blanks. PERSONNEL PSYCHOLOGY, *30*, 159–166.

Plummer JT. (1974). The concept and application of life style segmentation. *Journal of Marketing, 38*, 34.

Reilly RR, Chao GT. (1982). Validity and fairness of some alternative employee selection procedures. PERSONNEL PSYCHOLOGY, *35*, 1–62.

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

Russell CJ, Kuhnert KW. (1992). New frontiers in management selection systems: Where measurement technologies and theory collide. *Leadership Quarterly, 3*, 109–135.

Russell CJ, Mattson J, Devlin SE, Atwater D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology, 75*, 569–580.

Schoenfeldt LF. (1989, August). *Biographical data as the new frontier in employee selection research*. Address presented at the Annual Meeting of Division 5 of the American Psychological Association, New Orleans, LA.

Shaffer GS, Saunders V, Owens WA. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. PERSONNEL PSYCHOLOGY, *39*, 791–809.

Silverhart TA. (1990). *Career profile Australia/New Zealand validity report* (Technical report MRR4-1990). Hartford, CT: LIMRA.

Stricker LJ. (1988, November). *Assessing leadership potential at the Naval Academy with a biographical measure*. Paper presented at the Annual Meeting of the Military Testing Association, San Antonio, TX.

Turner JC. (1987). A self-categorization theory. In Turner JC (Ed.), *Rediscovering the social group* (pp. 42–67). New York: Basil Blackwell.

White LA, Kilcullen RN. (1992, August). *The validity of rational biodata scales*. Paper at the Annual Meeting of the American Psychological Association, Washington, DC.

White LA, Nord RD, Mael FA, Young M. (1993). The Assessment of Background and Life Experiences (ABLE). In Trent T, Lawrence J (Eds.), *Adaptability screening for the services* (pp. 101–162). Washington, DC: Office of the Asst. Secretary for Defense.