

Predictive Validity of Biodata Items Generated From Retrospective Life Experience Essays

Craig J. Russell

Institute of Management and Labor Relations Rutgers University

Joyce Mattson

Psychological Services Inc.

Steven E. Devlin

Navy Personnel Research and Development Center

David Atwater

ABSTRACT

Biographical information was generated from retrospective life-history essays completed in the fall of 1986 by first-year students at the U.S. Naval Academy. Essays were targeted to aspects of four generic life experiences (individual accomplishments, group accomplishments, disappointing situations, and stressful situations) that might reflect the processes or outcomes of antecedent developmental episodes. Essays were coded to generate biodata items, which were then given to 917 midshipmen entering the Naval Academy in the summer of 1987. Scales were developed that predicted criteria of military performance, academic performance, and peer ratings of leadership in validation and cross-validation samples. Implications for item development and biodata construct validity are discussed.

This research was supported, in part, by the U.S. Navy—American Society for Engineering Education Summer Faculty Research Program, the Navy Personnel Research and Development Center, and the Battelle Institute under Contract DAAG29-81-D-0100. The opinions expressed in this article are those of the authors, are not official, and do not necessarily reflect the views of the Navy Department.

We would like to thank Philip Bobko and three anonymous reviewers for their comments on earlier drafts; however, all responsibility for any errors in the content remains with us.

Correspondence may be addressed to Craig J. Russell, Krannert Graduate School of Management, Purdue University, Krannert Building, West Lafayette, Indiana, 47907.

Biographical information, or biodata, has been shown to capture systematic, enduring differences between subgroups of people. Biodata instruments have consistently demonstrated the ability to predict such criteria as job choice, vocational preferences, job turnover, and job performance (cf. [Brown, 1978](#); [Hunter & Hunter, 1984](#); [Mumford & Owens, 1982](#); [Owens & Schoenfeldt, 1979](#); [Schoenfeldt, 1989](#)). Biodata measures also tend to yield consistent factor structures over time and across samples (cf. [Davis, 1984](#); [Eberhardt & Muchinsky, 1982a](#), [1982b](#), [1984](#); [Neiner & Owens, 1982](#), [1985](#)).

Unfortunately, numerous researchers fail to provide any information on how their items were derived when they publish their results (e.g., [Brown, 1978](#); [Ritchie & Boehm, 1977](#)). Methods of item generation that have been reported implicitly involve two sets of constructs. The *target* constructs make up the criteria domain and typically consist of job-related knowledge, skills, and abilities; job performance measures; or job or career choices. The antecedent *developmental* constructs reflect the processes that precede performance on the targeted criterion constructs. Examples of the developmental construct domain include cognitions, affective responses, attitudes, behaviors, and demographic characteristics that result from previous life experiences.

For example, [Owens and Schoenfeldt \(1979\)](#) generated elements of developmental processes in their evaluation of the developmental—integrative (D—I) model ([Owens, 1968](#), [1971](#), [1976](#)). The D—I model is a general descriptive and predictive approach to human development that can be used to distinguish people with homogeneous life histories (including career choice). According to [Owens and Schoenfeldt \(1979\)](#), "fifty-two pages of topics for potential outcomes were developed in 'fleshing out' the outlines implied under input variables and prior behaviors. Two thousand items were constructed from or adapted to this outline" (p. 574). In other words, their research team used the D—I model as an outline of the developmental construct domain in hopes of creating construct-valid items. The actual items were created from investigators' best judgments (i.e., "fleshing out") regarding the substance of the developmental construct domain. Subsequent research indicates that this approach does yield items with consistent and meaningful underlying factor structures ([Mumford & Owens, 1984](#), [1987](#); [Mumford, Stokes, Owens, & Sparks, in press](#)).

Like the predictor-based efforts of Owens and his colleagues, criterion-driven methods for item generation have resulted in biodata scales that predict performance criteria in organizational settings. Investigators construct hypotheses about what kinds of developmental episodes are related to subsequent criterion performance. As noted by [Mumford and Owens \(1987\)](#),

"this process requires substantial judgment" (p. 5) on the part of investigators. Mumford and Owens suggested six sources of information for investigators to use in making these judgments: (a) the developmental literature; (b) life-history interviews with incumbents; (c) known life-history correlates of various characteristics; (d) typical factor loadings of background data items; (e) known predictive characteristics of various background data items; and (f) hypotheses formulated on the basis of general psychological knowledge. Notice that theories of human development can be used whether efforts to generate biodata items start with the criterion construct or the predictor construct domain. Furthermore, both approaches to item development rely heavily on the armchair judgments of investigators (cf. [Dunnette, 1968](#)). Nevertheless, the ability of a pool of biodata items to accurately reflect prior developmental constructs is dependent on the investigator's imagination or the validity and operational guidelines provided by some theory of human development or both.

Recent qualitative and theoretical efforts are just beginning to provide guidance for the identification of life-history events in the developmental construct domain that affect the subsequent performance of managers and leaders ([Lindsey, Homes, & McCall, 1987](#) ; [Kuhnert & Lewis, 1987](#) ; [Kuhnert & Russell, 1990](#)). Identifying life-history events that predict performance criteria is made easier when there is a great deal of overlap between the criterion construct domain and the developmental construct domain (i.e., it is presumably easier to identify developmental life episodes if candidates have had opportunities to show what they can do in previous positions). Criterion content sampling procedures will not work when (a) no previous opportunities existed (e.g., for entry-level jobs or for economically disadvantaged subgroups) or (b) entry-level personnel are selected with an ultimate criterion of performance in much higher, nonoverlapping jobs (e.g., the selection of "fast track" entry-level managers to be groomed for top-level positions).

These conditions face most professional schools and colleges attempting to select and subsequently educate high school seniors who are expected to perform well in managerial and leadership roles many years into the future. Alternative methods of generating pools of biodata items are needed if researchers are to obtain (a) greater insight into the underlying developmental construct domain and (b) enhanced predictive power when biodata is used in combination with cognitive measures currently employed. The purposes of this study were (a) to describe an alternative method of biodata item generation that captures meaningful life experiences not directly comparable to the job of U.S. Naval officer (i.e., high school experiences) and (b) to determine whether the biodata questionnaire developed in this way shows promise for predicting preliminary performance criteria.

Calls for thorough specification of job performance criteria as a guide for developing selection systems have characterized industrial/organizational psychology since its inception (cf. [Smith, 1976](#)). [Mitchell \(1986\)](#) and [Mumford](#)

[and Owens \(1987\)](#) have described general approaches to the content sampling of criterion construct domains for biodata item generation. In the present study, we attempted to decompose the target performance constructs for the position of naval officer and then generate examples of experiences in the learning or maturation process that reflect important antecedent developmental constructs.

Method

Sample

We used two samples for different stages of the study. Seventy first-year students at the U.S. Naval Academy (chosen at random) completed a set of autobiographical life-history essays in the spring of 1986. Twenty-eight of these students (again, randomly chosen) were subsequently interviewed in the fall of 1986 to clarify and expand upon the essay responses. None of the 70 students participated in any other part of the study.

Students from the U.S. Naval Academy Class of 1991 ($N = 917$; 836 men and 81 women; approximately 75% of the first-year students that completed the fall semester of 1987) participated in the second phase of the study. The biodata instrument was mailed to incoming students in May 1987. Candidates returned completed questionnaires by mail before arriving at the Naval Academy. The study is limited to the extent that responses of students already admitted to the Academy might differ from responses of students in the applicant pool (see [Mattson & Abrahams, in press](#), for an example of how this affects other biodata instruments in the Navy).

Development of Biodata Instrument

The criteria for success as a naval officer came from the Navy's officer-fitness report, a performance-appraisal instrument for officers. The performance dimensions used for instrument development were working relations, leadership, resource management, goal setting and achievement, and stress response. Other performance dimensions from the officer-fitness report were not used because high school students obviously lacked experience with certain military activities (i.e., "watch standing"). Each performance dimension was defined with a number of behavioral examples (e.g., "defines realistic goals," "develops plans and priorities," "responds positively to changing circumstances," and "involves subordinates in planning" for the goal setting and achievement dimension). Although the officer-fitness report may have certain problems in its day-to-day usage as a performance-appraisal device ([Cleveland, Morrison, & Burke, 1986](#)), we assumed that the performance dimensions were content-valid representations of what is expected of an officer in the Navy.

A pool of life experiences that might be related to these dimensions was generated from biographical essays written by the first sample of first-year

students ($N = 70$; see [Siegal, 1956](#), for a description of an early use of essays for this purpose). The essay topics were chosen so that antecedent developmental elements of the five key dimensions of officer fitness (described in the preceding paragraph) were likely to be reported. We content analyzed the biographical essays to generate biodata items for the self-report biodata questionnaire. [Appendix A](#) contains the essays and the content analysis framework used to generate items from the 70 midshipmen's narrative responses. The actual procedures for item generation were as follows.

In the spring of 1986, we distributed the biographical essay assignment to 70 first-year midshipmen, who completed the essays and returned them in early October 1986.

We content analyzed the essays, using the framework in [Appendix A](#). Using summary coding sheets, we abstracted inputs (e.g., parental support), behavioral outcomes (e.g., working one-on-one with an adult), affective responses (e.g., feeling cheated out of something), and perceptions or cognitions (e.g., thinking some group member was not pulling his or her weight) from the essays. Biodata items were developed from each life experience extracted from the essays; however, the incidents abstracted tended not to have a one-to-one correspondence with the content analysis framework. For example, one student described a stressful situation in which he had been going steady with three different girls during his senior year in high school. He was able to do this because the girls attended different schools in the large metropolitan area where he lived. At prom time, however, he had to decide who to take and how to communicate this to the three girls. The student's description of this incident was used to generate the following items:

- Many people date others on a steady basis in high school. How often have you had to "break up" with someone you had been seeing on a steady basis?
- How often have you lost sleep because you were anxious about something in high school?
- How often has someone fooled you, so that afterwards you felt stupid or naive?

The first item was generated as an example of "display integrity and moral courage" from the content analysis examples of leadership listed for the stress situation essay in [Appendix A](#), but the other two items did not correspond with any headings found in the content analysis form. Consequently, although we used the content analysis form as a starting point to ensure some congruence between the criterion domain and the developmental constructs reflected in the biodata items, approximately half of the items had no obvious relationship. At this point there is no reason to believe that any one way of extracting biodata items from-life history essays is more appropriate than another. The purpose of the

coding process was to extract every possible life experience present in the essays, *not* to identify or extrapolate items that rigidly fit some a priori framework.

In most cases, two items were generated for each incident. The first item asked how frequently this type of situation occurred; the second asked how often this type of situation was handled effectively. This strategy corresponds with the distinction drawn by [Mumford, Stokes, Owens, and Sparks \(in press\)](#) between exposure (how often did it happen) and adaptation (what happened during the exposure). Each response scale consisted of a 5-point continuum ranging from *never* (1) to *very often* (5). For some incidents, a response-effectiveness dimension was omitted, either because it was not relevant (e.g., parental divorce or death of a peer) or because it could not be assessed with a frequency-response format (e.g., "How good were your study skills in high school?"). [Kane \(1987\)](#) described this item-generation procedure for performance-appraisal applications. (See [Russell, 1986](#), for the complete set of items and examples of completed essays.)

In late October 1986, we interviewed 28 of the midshipmen who had completed the biographical essays to ensure that the biodata items generated accurately represented their essay responses. Each interview lasted 30 min. With the midshipmen's permission, all interviews were tape recorded, transcribed, and used with the original biographical essays in item development. We then placed items in a questionnaire format for administration to entering midshipmen in the spring of 1987.

Performance Criteria

Although performance as a Naval officer was the ultimate criterion of interest in this research, such a measure is obviously not available for individuals who have just entered the Naval Academy. For this reason, five surrogate criteria tapping performance at the Naval Academy were used for the preliminary validity analyses. These criteria included an academic quality point ratio (AQPR), a military quality point rating (MQPR), a military performance rating, a conduct rating, and a peer rating of leadership. The AQPR was the student's academic grade point average at the end of the first semester at the Naval Academy. AQPR is a continuous variable ranging from 0 (low) to 4 (high).

The MQPR was a global measure of military performance, derived at the end of the first semester from a weighted combination of physical education grades (weight = 2), military performance ratings (weight = 6), conduct ratings (weight = 3), and grades in professional courses such as naval engineering and fundamentals of naval science (weight = $1.72 \times \text{number of credit hours}$). The weights are assigned by the Naval Academy administration. Because class grades are included, the MQPR must be viewed as a contaminated measure of managerial or leadership performance. The military performance and conduct components of the MQPR closely parallel the Navy's officer-fitness report and

were split out for separate analyses. MQPR is a continuous variable ranging from 0 (low) to 4 (high).

The military performance rating consisted of a grade assigned by the student's company officer on the basis of performance assessments made by peers and senior officers. As input to this process, each student rates the two peers she or he believes rank highest in the company and the two peers she or he believes rank lowest. Others in a position of authority over a midshipman who have an opportunity to observe him or her also provide input. Both the peer and superior officers' input is gathered on a form very similar to the form used for rating officer performance in the fleet. One purpose of this rating procedure is to familiarize students with the performance-appraisal process they will encounter after graduation. The company officer for each midshipman subjectively arrives at a final military performance rating by combining these evaluations with his or her own evaluation. The military performance rating is discrete and ranges from 0 (low) to 4 (high).

The conduct rating is based on the number of conduct demerits a student accrues. Demerits are received for a wide range of offenses, which are generally described as "conduct which might reflect discredit upon the Brigade of Midshipmen or the Naval Academy, or is in violation of any local, state or federal law, or indicates questionable personal morals" ([U.S. Naval Academy, 1986](#), p. 1.0—1). The conduct grade is derived from a table of demerit ranges associated with each grade for a given class. For example, a first-year student with 0 to 10 demerits might receive a conduct rating of 4, a first-year student with 11 to 20 demerits, a rating of 3, and so forth. The ranges get smaller for sophomores, juniors, and seniors. The actual number of demerits received by a student is not centrally recorded and was unavailable to us. Conduct ratings are discrete and range from 0 to 4, with 4 being high.

The peer ratings of leadership were not part of the normal Naval Academy performance-evaluation system. These ratings were collected by [Stricker \(1988\)](#) at the end of summer orientation and before the beginning of the first year (i.e., at the end of the military-activities-oriented summer program before students began their fall academic curriculum). The ratings were not part of any institutionally mandated performance measure. For this purpose, Stricker selected half of the squads of first-year midshipmen and had members of those squads anonymously rate themselves and other members of their squad on the dimension of leadership. A 9-point scale was used, with ratings ranging from *extremely able to lead* (9) to *extremely unable to lead* (1). An average of 11 peer ratings were collected for each individual. The median rating received by each student within his or her squad served as the leadership criterion. Ratings were evaluated for interrater agreement by correlating median ratings for one half of the squad's evaluations with median ratings for the other half. Interrater agreement calculated in this way, using the Spearman-Brown correction for double length, was .89.

The means, standard deviations, and intercorrelations of all of the criteria are shown in [Table 1](#). Examination of the frequency distributions indicates that the distribution of conduct ratings was negatively skewed and may not provide adequate variance for empirical scale construction.

Scale Construction

We constructed separate biodata scales for each criterion by randomly selecting two thirds ($N = 563$) of the men who were first-year students in the fall of 1987. Only men were selected for this purpose because (a) biodata items often demonstrate different predictive relationships and internal factor structures for men and women (cf. [Owens & Schoenfeldt, 1979](#)) and (b) there were not enough women for separate scale construction.

We used a contrasting-groups approach to key development ([England, 1961](#)) to form each scale. Separate predictor scales were developed by contrasting all biodata responses of individuals in the top third of their respective criterion distributions with responses of those in the bottom third. Hence, separate scales were developed for the AQPR, MQPR and leadership criteria. Predictor scales for the military performance and conduct ratings were developed differently because of the uneven distribution of their discrete ratings. To construct the scale predicting military performance, we contrasted individuals receiving grades of 1 or 2 with those receiving grades of 4. For the conduct scale, we contrasted individuals receiving grades of 0 through 3 with individuals whose grades were 4. The skewed distributions of the military performance and conduct grades were admittedly suboptimal for scale construction, but attempts were nevertheless made to develop scales. For each scale, the 50 most discriminating item responses were selected to be scored. The minimum scored difference in item-response percentages between high and low criterion groups ranged from 7% for the AQPR scale to 9.7% for the leadership scale.

Weights of +1 were applied to scored item responses that the high-criterion group endorsed more often than the low-criterion group, and weights of - 1 were assigned to item responses for which the reverse was true. All scales were cross-validated on a sample consisting of the remaining one third of the men ($n = 273$) and on another sample consisting of all of the women ($n = 81$). The potential usefulness of each scale in terms of its ability to increment the predictive power of the present Naval Academy selection composite (known as the Candidate Multiple; CM) also was evaluated with the hold-out sample of 273 men. The CM is an empirically developed composite in which Scholastic Aptitude Test scores, high-school performance, extracurricular activities, letters of recommendation, and vocational interest scales are weighted to predict academic performance, military performance, academic major, and disenrollment.

We factor analyzed items scored on one or more of the cross-validated biodata scales (i.e., all except the Conduct scale) to identify any underlying developmental constructs. Such factors could then be tapped more thoroughly in future revisions of the biodata instrument. We selected 71 items by developing a revised scale for each criterion based on the total sample of men (many items were predictive of more than one criterion, so that a total of 71 items demonstrated predictive validity for at least one criterion). All 100 items were originally entered into a factor analysis. Unfortunately, regardless of the algorithm used, the computer could not arrive at a solution. Error messages indicated that there was a perfect correlation between some linear combination of the items and another linear combination of the items. This relationship did not occur for the smaller subset of 71 items found to be predictive of the various performance criteria. An initial risk was incurred of capturing only those aspects of the life-history construct domain that demonstrated criterion-related validity in the current study.

Nevertheless, a matrix of Pearson product—moment correlation coefficients was then generated. Communality values based on iterative estimates of the squared multiple correlation of each variable with the remaining variables in the matrix were inserted into the diagonal of the matrix, and factors were extracted by using the principal axis method with oblique rotations. Factor pattern matrices for four- and five-factor solutions were examined to determine if item loadings might correspond with the four sets of stimulus materials (i.e., group accomplishment, individual accomplishment, disappointing situations, and stressful situations) or the five performance categories (working relationships, leadership, resource management, goal setting and achievement, and stress response) or both.

Results

Cross-validities of the five scales with the criteria for male students are presented in [Table 2](#). Many validities were not only significant but also of a magnitude suggesting the potential for practical utility. With the female sample, none of the biodata scales demonstrated significant cross-validities with their respective criteria. ¹

The incremental validity results in [Table 3](#) suggest that the biodata questionnaire has definite potential for improving the prediction of MQPR and military performance. However, because the CM was designed to predict a number of criteria simultaneously and is likely to be suboptimal for predicting any of the criteria individually, additional comparisons against an optimal CM composite for each criterion were needed. Accordingly, selection variables that were included in the CM were optimally weighted to predict each of the individual criteria in turn, and incremental validities relative to these optimal composites were computed. These analyses (not shown) yielded slightly reduced, but still significant, increments to the validities. However, because the CM was developed and used in actual candidate selection, it was subjected to direct range restriction, whereas

the biodata instrument was exposed to indirect range restriction. All students in the current study had been admitted to the Naval Academy (both the validation sample and the essay writers). Essays and responses to the biodata items might be very different for the 13,000 applicants than for the students who were admitted and made it through their first semester. Consequently, the reported incremental changes in criterion related validity must be viewed with caution.

Factor analysis was performed on the 71 items that demonstrated predictive validity for at least one criterion. The inter-item correlation matrix was based on the entire range of responses on each item (i.e., it was not limited to just those response options that contributed to the empirical key). It must be kept in mind that [England's \(1961\)](#) contrasting-group approach uses only those response options that differentiate between high and low performers in the key development sample. Total within-item variance is represented by the entire range of experience frequencies, whereas some subset of that variance (associated with *only* those response options that differentiate between high and low performers) represents that range of experience frequencies that predict subsequent performance. Factor analysis results reflect responses to the entire range of response options, not just those that contribute to prediction. To perform factor analysis on only those response options that differentiate between high and low performers would require (a) that each response option (1 through 5) be viewed as an item unto itself and (b) factor analysis of a biserial or tetrachoric correlation matrix among each of these dummy-coded response options. Besides having undesirable psychometric characteristics associated with the correlation matrix computed among dichotomous dummy-coded response options, any factor structure would be extremely difficult to interpret. As most of the original items contributed more than one response option to the scoring key, one might expect 71 interpretable "factors" representing each of the original 71 items. Consequently, although the factor analysis was limited to only the 71 items that contributed to the empirical key, total within-item variance was factored, not just the portion of variance contributed by the keyed response options.

Common factor analysis performed on the entire range of response options generated five interpretable factors accounting for 21.3% of the total variance, not an uncommonly low amount of variance for the biodata literature ([Owens, 1976](#)). Items loading dominantly on factors derived in the four-factor solution continued to load dominantly on the first four factors of the five-factor solution (2 items that had not received a dominant loading in the four-factor solution generated dominant loadings on Factor 5). The items loading on each factor are presented in [Appendix B](#) . Subjective content interpretations resulted in preliminary factor labels of Life Problems and Difficulties, Aspects of Task Performance, Work Ethic/Self-Discipline, Assistance From Others, and Extraordinary Goals or Effort.

Finally, to examine internal consistency reliability of items loading on a single factor and of factor intercorrelations, we calculated scale scores for each factor.

Response options contributing to the scoring key for any item loading higher than .30 on a factor were dummy coded and examined for internal consistency reliability. The means, standard deviations, coefficient alphas, and intercorrelations of the *empirically keyed* scale scores are presented in [Table 4](#) (note that coefficient alphas are calculated for each of the criterion-based keys, and hence the range of alphas is reported in [Table 4](#)). Unfortunately, empirical keys based on contrasting-groups procedures suffer from severe range restriction when the number of items on a scale is 10 or fewer. The last five columns of [Table 4](#) contain cross-validities between the five criterion measures and the empirically keyed scale scores for the five-factor solution. Regardless, the primary purpose of the factor analysis was to delineate components of the developmental construct domain for future item development.

Discussion

Empirically derived biodata scales predicted peer ratings of leadership as well as ratings of military and academic performance. Furthermore, the biodata scales significantly contributed to the prediction of military and academic performance beyond validities generated by the current predictor set; the biodata items developed by using the retrospective life-experience method appear worthy of further study for predicting Naval Academy performance and subsequent performance as a Naval officer. By extension, the method holds promise for prediction of performance in other varieties of professional schools.

Regardless of any theoretical implications for the developmental construct domain, implications for practice are immediate. Currently, the practitioner who wants to use a biodata-based selection system must either go through the laborious and vague procedures needed to generate a new item pool (cf. [Mumford & Owens, 1987](#); [Owens & Schoenfeldt, 1979](#)) or purchase one of the proprietary instruments available from vendors (cf. [Sharf, 1981](#)). After obtaining some finite number of items, the user is still faced with security problems. Life-history essays and interviews can be used to generate large item pools (with a high percentage of criterion-valid items) relatively quickly and inexpensively. Indeed, [Russell and Domm \(1990\)](#) recently generated 550 items from life-history essays written about experiences reflecting assessment-center dimensions. The entire procedure took two developers approximately 4 days over a 4-week period. Ease in item generation permits the development of multiple equivalent forms and the constant updating of item banks, qualities not found in most current biodata applications.

The life-history procedure also has implications for theory development. The theoretical goal of biodata research is to develop some model of human development that specifies operational procedures resulting in valid measures. These measures should capture the salient developmental constructs and demonstrate criterion-related validity. Procedures used in the current study were only weakly specified by [Owens's \(1968, 1971\)](#) D—I model of human

development. We used stimulus materials (essays) which assumed that individual and group accomplishments, disappointing situations, and stressful situations would capture important aspects of the developmental construct domain. The coding procedures used to extract biographical incidents were purposefully vague to avoid loss of information. Both the essays and the coding materials also reflected the criterion construct domain in hopes that any developmental constructs captured by life experiences described in the essays would be related to the criterion of interest. Consequently, one can only speculate whether the resultant life-history incidents represent true developmental episodes that causally affect subsequent job performance, or incidental fall-out (and hence only correlates) of the developmental process, or both.

The factor analytic results lend themselves to just such speculation. With the possible exception of Factor 2 (Aspects of Task Performance), the factors did not have simple correspondence with the criterion construct domain. For example, items loading on Factor 1 suggest that a dominant developmental construct might involve aspects of dealing with life's problems. Recent results reported by [Lindsey et al. \(1987\)](#) in their study of key events in managers' lives lend support to this interpretation. Specifically, Lindsey et al. reported that one of the experiences most often cited as a key developmental episode by top-level executives was having to work for a demanding, stress-inducing boss. It appears that having to deal with life's problems is a richly recalled experience at initial and later career stages. However, it is still unclear whether items loading on Factor 1 represent (a) one set of basic developmental building blocks from which high-school students extract meaning or (b) correlates or incidental fallout that predict but do not explain subsequent job performance. [Kuhnert and Russell \(1990\)](#) used a constructive/developmental model of human development to generate biodata items corresponding to the different ways in which an individual might "make meaning" out of a life experience. Whether this process is called development, learning, or meaning making, biodata will remain an old technology on the verge of being a new frontier in personnel selection until efforts are directed at determining how life-experiences affect the individual ([Fleishman, 1988](#) ; [Schoenfeldt, 1989](#)).

Results of the factor analysis also indicate a number of directions for development of future essay stimulus material and coding procedures, which might shed further light on underlying constructs. For example, essay questions about having to work hard or exert self-discipline should generate more biographical incidents that would load on Factor 3, enhancing its internal consistency reliability by adding more items. In addition to using a content-analysis framework derived from the targeted criterion domain, future essay coders could independently use a framework derived from the factor structure. One group of coders could use a coding framework taken from the criterion domain while another set of coders used a preliminary framework of the developmental construct domain derived from the factor analysis. Finally,

Factors, 1, 2, and 3 yielded different patterns of cross-validities for the five criteria. Although the current set of criteria are admittedly surrogates for the ultimate criterion of performance as an officer in the fleet, different patterns of cross-validity results imply that different types of coding emphasis or essay stimulus materials might generate valid items for some criteria and not others. For example, in the current study, items loading on Factors 1 and 3 were highly correlated with one another ($r = .53$) and with the MQPR. Interpretations of common themes underlying these correlated factors may facilitate the construction of new essays that should yield items with higher predictive power (e.g., one common theme may involve seeking assistance from others for help with life's problems).

Finally, we used components of the criterion construct domain to create essay questions and coding materials. Future researchers need to examine how theories of human development can be used to create essay stimulus materials and subsequent coding procedures. For example, [Mumford and Owens \(1984\)](#), [Mumford, Stokes, and Owens \(in press\)](#), and [Mumford, Stokes, Owens, and Sparks \(in press\)](#) have extended Owens's original D—I model ([Owens, 1968, 1971](#)) to an ecology model, the most detailed model of human development in the biodata literature. Critical components of this model include notions of exposure and adaptation processes associated with various life experiences.

Future essay and interview procedures might be structured to generate examples of the frequency of exposure to various life situations (e.g., individual accomplishments, part-time employment in high school, extracurricular activities). This is the most common type of biodata item currently in use, though more work needs to be directed at the cognitive aspects of self-assessment ([Ashford, 1989](#)). Procedures might also be structured to yield examples of components of the adaptive processes that occur in each life situation (e.g., perceptual processes, behaviors, attitudes, values, beliefs, situation outcomes). In addition, essays and interviews could be designed to generate examples of relationships between life experiences and adaptive processes over time. If homogeneous patterns of life experiences and adaptive processes exist, major strides can be made in both selection from and development of applicant pools.

Other investigators have gathered life-experience information in a much less systematic manner. [Lindsey et al. \(1987\)](#) and [Russell \(1990\)](#) reported the use of structured interview procedures to gather extremely rich life-history information on top-level executives. [Hough \(1984\)](#) and [Love and O'Hara \(1987\)](#) have gathered critical life experiences using much less structured stimulus or coding materials or both. These investigators have focused on how to qualitatively interpret and subsequently scale these life experiences. The current results indicate that structured stimulus *and* coding materials can be used to cost-effectively generate criterion-valid biodata items. The relative ease with which large numbers of items can be generated holds great promise for both criterion prediction and theory development.

APPENDIX A

A Biographical Essays and Content Analysis Formats Essay 1

Most high school age individuals get involved in group efforts of one kind or another. These efforts can generally be described in terms of some desired goals or objective. Further, these efforts usually involve limited resources; you don't have all the money, time or "resources" in the world! Please describe a group effort you were involved in that was aimed at achieving some goal. Try to address the questions provided below in your description.

- What was the goal of the group (e.g. build a float for a parade, raise money for charity, earn money during summer vacation, etc.)? How was this goal decided upon? Were you particularly committed to this goal or was it just something to do?
- What role did you play in the group effort—what were you expected to do? How was it decided who should do what? Were you happy with your role—why or why not? Did anything in your past experience help you in this activity?
- What resources were available to aid your achievement of the goal? Was there a major limiting factor?
- What were your working relations with others? Did you like these relationships?
- Was the goal achieved?
- What did you learn from this experience?

Essay 2

Most people can point to some situation in high school where they accomplished something important. It wasn't necessarily something other people might all agree was important, but it was important to the person involved. It made him/her feel good. Further, the accomplishment was achieved mainly because of the individual's own efforts (though others might have been involved). Please describe this accomplishment in Essay 2. Try to answer all of the following questions in your description.

- What did you achieve, and why was it important to you? Did you feel you tried your best? What past experiences have you had that helped you achieve this accomplishment?
- Why did you try to achieve this particular goal? What, if any, other objectives did you consider?
- What alternative ways of achieving this goal did you consider? Why did you decide to do what you did?
- What obstacles did you have to overcome? Any surprises?
- What did you learn from this experience?

Essay 3

In addition to accomplishments, everybody experiences disappointments from time to time. Please describe your greatest disappointment while in high school. Try to answer the following questions in your description.

- Why did you decide to try and accomplish this particular goal?
- What obstacles did you identify beforehand that might have prevented you from achieving your goal? Were there any obstacles that you did not anticipate?
- What kinds of resources (money, time, adult help, etc.) did you have access to?
- Were there alternative ways of achieving the goal that you considered? Why did you choose the one you did?
- What did you learn from this experience?

Essay 4

From time to time most people find themselves in fairly stressful situations. Further, what might be stressful to one person is a piece of cake to another. Please describe the most stressful situation you have encountered in the last few years. Try to answer the following questions in your description.

- What was expected of you in this situation (if multiple things were expected of you by different people, please note all of them)? What did you expect of yourself?
- Was there anything in your prior experiences that helped prepare you for dealing with the stress?
- What was it about the situation that caused you the most stress? How did you try to manage it (did you think of anything or do anything in particular)? Did you consider doing anything else at the time?
- Would this situation cause you as much stress in the future? Why or why not?
- What did you learn from this situation?

Content analysis format

The following guidelines were used in reading Essays 1—4 to generate bioincidents from which the multiple-choice biodata items were developed.

Essay 1: Group Project

- Working Relations: look for examples of...
 - promote spirit of teamwork
 - give constructive feedback and help group members
 - encourage two-way communication among members and/or group leader

- help in matching of team members with different skills to the most appropriate part of the task
- Leadership: look for examples of...
 - proper use of authority in meeting responsibilities
 - instruction to subordinates where appropriate
 - recognition and use of subordinate strengths
 - contribution to unit's morale
 - consider and encourage ideas of subordinates
 - delegates authority appropriately
 - display integrity and moral courage
 - promote teamwork
- Resource Management: look for examples of...
 - demonstrates knowledge of operation, maintenance, and safety in use of equipment
 - knows relative costs of required resources
 - sets up systems to monitor scarce resource usage
 - fiscal authority/responsibility

Essay 2: Individual Achievement

- Goal Setting and Achievement: look for examples of...
 - pursuit of new learning experiences
 - goals set in timely fashion
 - setting of moderately difficult, specific goals
 - prioritize and develop plans
 - flexible response to changing circumstances
 - goal achievement
 - seeking of constructive feedback out of setbacks—use of feedback in setting up new strategies and/or goals
 - setting of realistic priorities among goals competing for his/her time and resources
 - efficient use of resources available to achieve goal
- Resource Management: look for examples of...
 - demonstration of knowledge of operation, maintenance, and safety in use of equipment
 - knowledge of relative cost of resources required
 - construction of control systems to ensure proper use of limited resources
 - fiscal authority/responsibility

Essay 3: Disappointment

- Goal Setting and Achievement: look for examples of...
 - seeking of constructive feedback to revise goals and/or strategies and tactics
 - setting of goals in timely fashion
 - setting of moderately difficult, specific goals
 - development of plans and priorities
 - flexible response to changing circumstances
 - realistic priorities among competing goals
 - efficient use of resources for goal attainment
- Resource Management: look for examples of...
 - knowledge of operation, maintenance, and safety in use of equipment
 - knowledge of relative cost of resources
 - setting up of control systems to monitor limited resources
 - fiscal authority/responsibilities
- Stress Response: look for examples of...
 - retention of composure and effectiveness
 - recognition of obstacles and gathering of needed information
 - decisive actions
 - generation of alternative responses for various contingencies

Essay 4: Stress Situation

- Leadership: look for examples of...
 - contribution to group's morale
 - consideration and encouragement of ideas of others
 - display of integrity and moral courage
 - promotion of spirit of teamwork
 - delegation of authority
 - use of subordinates' strengths to achieve goals
 - obtains or gives instruction when needed
- Resource Management: look for examples of...
 - knowledge of operation, maintenance, and safety in use of equipment
 - knowledge of relative cost of needed resources
 - fiscal authority/responsibility
- Stress Response: look for examples of...

Brown, S. H. (1978). Long-term validity of a personal history item scoring

procedure. *Journal of Applied Psychology*, 63, 673-676.

Cleveland, J. N., Morrison, R. F. & Burke, D. G. (1986, April). *Rater intentions in appraisal ratings: Malevolent manipulation or functional "fudging"?* (Presented at the First Annual Meeting of the Society for Industrial and Organizational Psychology, Chicago)

Davis, K. R. (1984). A longitudinal analysis of biographical sub-groups using Owen's developmental—integrative model. *Personnel Psychology*, 37, 1-14.

Dunnette, M. D. (1968). *Personnel selection and placement*. (Belmont, CA: Wadsworth)

Eberhardt, B. J. & Muchinsky, P. M. (1982a). An empirical investigation of the factor stability of Owen's biographical questionnaire. *Journal of Applied Psychology*, 67, 138-145.

Eberhardt, B. J. & Muchinsky, P. M. (1982b). Biodata determinants of vocational typology: An integration of two paradigms. *Journal of Applied Psychology*, 67, 714-727.

Eberhardt, B. J. & Muchinsky, P. M. (1984). Structural validation of Holland's hexagonal model: Vocational classification through the use of biodata. *Journal of Applied Psychology*, 69, 174-181.

England, G. W. (1961). *Development and use of weighted application blanks*. (Dubuque, IA: W. C. Brown)

Fleishman, E. A. (1988). Some new frontiers in personnel selection research. *Personnel Psychology*, 41, 679-701.

Hough, L. M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology*, 69, 135-146.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.

Kane, J. S. (1987). Performance distribution assessment. (In Ronald Berk (Ed.), *Performance assessment: Methods and applications* (pp. 100—120). Baltimore: Johns Hopkins University Press.)

Kuhnert, K. W. & Lewis, P. (1987). Transactional and transformational leadership: A constructive/developmental analysis. *Academy of Management Review*, 12, 648-657.

Kuhnert, K. W. & Russell, C. J. (1990). Using constructive developmental theory and biodata to bridge the gap between personnel selection and leadership. *Journal of Management*, 16, (3) 1-13.

Lindsey, E., Homes, Z. & McCall, M. (1987). *Key events in executive lives*. (Greensboro, NC: Center for Creative Leadership)

Love, K. G. & O'Hara, K. (1987). Predicting job performance of youth trainees under a job training partnership act program (JTPA): Criterion validation of a behavior-based measure of work maturity. *Personnel Psychology*, 40, 323-340.

Mattson, J. & Abrahams, N. E. (in press). *Preliminary evaluation of a personal history questionnaire for predicting Naval Academy disenrollment*. (San Diego, CA: Navy Personnel Research and Development Center)

Mitchell, T. W. (1986, April). *Specialized job analysis for developing rationally oriented biodata prediction systems*. (Paper presented at the First Annual

Meeting of the Society of Industrial and Organizational Psychology, Chicago)

Mumford, M. D. & Owens, W. A. (1982). Life history and vocational interests. *Journal of Vocational Behavior*, 21, 330-348.

Mumford, M. D. & Owens, W. A. (1984). Individuality in a developmental context: Some empirical and theoretical considerations. *Human Development*, 27, 84-108.

Mumford, M. D. & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement*, 2, 1-31.

Mumford, M. D., Stokes, G. S. & Owens, W. A. (in press). *Patterns of life adaptation: The ecology of human individuality*. (Hillsdale, NJ: Erlbaum)

Mumford, M. D., Stokes, G. S., Owens, W. A. & Sparks, C. P. (in press). Developmental determinants of individual action: Theory and practice in the application of background data measures. (In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (2nd ed.). Orlando, FL: Consulting Psychologist Press.)

Neiner, A. G. & Owens, W. A. (1982). Relationships between two sets of biodata with 7 years separation. *Journal of Applied Psychology*, 67, 146-150.

Neiner, A. G. & Owens, W. A. (1985). Using biodata to predict job choice among college graduates. *Journal of Applied Psychology*, 70, 127-136.

Owens, W. A. (1968). Toward one discipline of scientific psychology. *American Psychologist*, 23, 782-785.

Owens, W. A. (1971). A quasi-actuarial prospect for individual assessment. *American Psychologist*, 26, 992-999.

Owens, W. A. (1976). Background data. (In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 609—644). Chicago: Rand McNally.)

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

Ritchie, R. J. & Boehm, V. R. (1977). Biographical data as a predictor of women's and men's management potential. *Journal of Vocational Behavior*, 11, 363-368.

Russell, C. J. (1986). *Review of the literature and development of a biodata instrument for prediction of naval officer performance at the U.S. Naval Academy* (Contract No. DAAG29-81-D-0100). (San Diego, CA: Navy Personnel Research and Development Center)

Russell, C. J. (1990). Selecting top corporate leaders: An example of biographical information. *Journal of Management*, 16, 71-84.

Russell, C. J. & Domm, D. R. (1990, April). *On the construct validity of biographical information: Evaluation of a theory-based model of item generation*. (Presented at the Fifth Annual Meeting of the Society of Industrial and Organizational Psychology, Miami Beach, FL)

Schoenfeldt, L. F. (1989, August). *Biographical data as the new frontier of employee selection research*. (Paper presented at the 97th Annual Convention of the American Psychological Association, New Orleans, LA)

Sharf, J. (1981). *Supervisory profile record technical reports* (Vol. 1—3). (Washington, DC: Richardson, Bellows, Henry)

Siegal, L. (1956). A biographical inventory for students: I. Construction and standardization of the instrument. *Journal of Applied Psychology*, 40, 5-10.

Smith, P. C. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria.(In M. D. Dunnette (Ed.), *The handbook of industrial and organizational psychology* (pp. 745—775) Chicago: Rand McNally.)

Stricker, L. (1988). *Assessing leadership potential at the Naval Academy with a biographical measure* (Contract No. DAAL03-86-D-001).(San Diego, CA: Navy Personel Research and Development Center)

U.S. Naval Academy (1986, July). *Operation and administration of military performance system* (Commandant of Midshipmen Instruction 1610.6f).(Annapolis, MD: Author)

1

One of the 25 validity coefficients derived for the female subsample was significantly different from zero at $p < .05$. This is not surprising, given the results reported by [Owens and Schoenfeldt \(1979\)](#). Results for parallel analyses on the female subsample can be obtained from Craig J. Russell.

Table 1.

Table 1
Means, Standard Deviations, and Intercorrelations of Criteria

Criteria	Range	M	SD	Intercorrelations				
				1	2	3	4	5
1. Leadership	0-4	0.12	0.71	—	.07	.41	.45	.17
2. AQPR	0-4	1.15	0.42		—	.70	.42	.17
3. MQPR	0-4	1.80	0.72			—	.60	.46
4. Military performance	0-4	1.05	0.66				—	.39
5. Conduct	0-4	1.81	0.54					—

Note. N = 836 men except for the leadership criterion, for which N = 455 men. High ratings represent the most desirable behavior for each criterion. AQPR = academic quality point rating; MQPR = military quality point rating.

Table 2.

Table 2
Cross-Validities of Biographical Questionnaire Scales

Scale	Criteria				
	1	2	3	4	5
1. Leadership	.19*	-.02	.06	.30	-.02
2. AQPR	-.03	.20**	-.20**	-.27**	.17**
3. MQPR	.03	.13**	.30**	.20**	.13**
4. Military performance	.09	.04	.20**	.19**	.10*
5. Conduct	.00	.16**	.11*	.12*	.01

Note. N = 273 men except for the leadership criterion, for which N = 136 men. Each biodata score comprised 50 items. AQPR = academic quality point rating; MQPR = military quality point rating.

* $p < .05$. ** $p < .01$.

Table 3.

Table 3
Incremental Validation of Biographical Questionnaire Scales

Predictor	Criterion			
	Leadership	MQPM	MQPM	Military performance
CM				
CM + Leadership scale	.16	.02 ^a	.04 ^a	.02 ^a
R^2	.03			
CM + MQPM scale	.18			
R^2	.04 ^a			
CM + MQPM scale	.18			
R^2	.04 ^a			
CM + Military performance		.04 ^a		
R^2		.02		
CM + Leadership scale			.04 ^a	
R^2			.02	
CM + Leadership scale				.03
R^2				.01

Note. N = 273 men except for the leadership criterion, for which N = 136 men. MQPM = military quality performance scale; CM = combatant; CM + CM = combatant sample size. CM, combat leadership's selection component.
^ap < .05. ^bp < .01.

Table 4.

Table 4
Data for Item, Item-to-Item, Item-to-Item, Item-to-Item, and Item-to-Item in Biographical Questionnaire Scales

Item	Item-to-Item									
	Item	Item	Item	Item	Item	Item	Item	Item	Item	Item
Item 1	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 2	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 3	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 4	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 5	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 6	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 7	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 8	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 9	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Item 10	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10

Note. N = 136 men except for the leadership criterion, for which N = 273 men. MQPM = military quality performance scale; CM = combatant; CM + CM = combatant sample size. CM, combat leadership's selection component.
^ap < .05. ^bp < .01.