CHAPTER

2

Generation Procedures for Biodata Items
A Point of Departure

Craig J. Russell
Louisana State University

The purpose of this chapter is to help future investigators generate biodata items that demonstrate both construct- and criterion-related validity. It is my belief that such efforts will ultimately lead to the development and evaluation of theories of performance prediction (see Campbell, 1990). The first part of this chapter is devoted to convincing the reader of the merits of such an effort. Theories of life history events, such as they are, are described in light of the absence of any operational guidelines for biodata item generation. I argue that the gap between theories of life history and operationalizations (i.e., biodata item content) presents a major opportunity for scholarly contribution to a theory of performance prediction, or, as borrowed from Fleishman's (1988) discussion of similar issues, a "new frontier."

The second part of this chapter provides explicit examples of two item generation procedures applied to three theories of individual differences (personality, vocational interest, and leadership). This section portrays a number of item generation efforts I have been involved in over the past decade. On my first involvement with biodata, I found the absence of published instruments and keys to be a major barrier to entry. I hope this section will give the novice biodata researcher a number of alternate points of departure.

The third section describes how item content might affect other aspects of biodata items. These include item reliability, heterogeneity, behavioral discreteness, and behavioral consistency.

IMPORTANCE OF BIODATA ITEM CONTENT

Recent research efforts involving biographical information typically start with a brief overview of meta-analytic results supporting biodata criterion-related validity (e.g., Hunter & Hunter, 1984; Reilly & Chao, 1982; Schmitt, Gooding, Noe, & Kirsch, 1984). Biodata research conducted prior to the introduction of meta-analytic techniques usually referenced Ghiselli's (1966) survey of test validities. Each biodata study would then describe the particular aspect or issue to be examined. Recently, these have included examinations of alternate scoring procedures (Mitchell, 1992; Mitchell & Klimoski, 1982), the susceptibility of biodata items to faking in applicant versus incumbent samples (Hogan & Stokes, 1989; Kluger, Reilly, & Russell, 1991), the effect on criterion-related validity of biodata response distortions (Trent, 1987), the generalizability of item validities across jobs (Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990), and the consistency of biodata factor structures over time (Neiner & Owens, 1982).

However, until Mumford and Owens (1987) published their review of methods in biodata technology, guidelines on how to generate items were not available in the literature. Few investigators provided any direction on how to construct an item that could be expected to demonstrate criterion-related validity. Almost no investigators have described explicitly how to develop items that could subsequently be used to test theory-based hypotheses (for an exception, see Kuhnert & Russell's [1990] description of how Kegan's [1982] constructive/developmental theory of adult development might be used to generate biodata items).

That is not to suggest that efforts at developing models or theories of *constructs* underlying biodata items have not taken place. Owens' (1968, 1971) developmental-integrative model suggests that prior life events are sources of individual development and integration (i.e., the meaning a person derives from an event) that influence future knowledge, skills, abilities, and motivation. Mumford and Stokes (1991) have extended this approach to an ecology model describing a longitudinal sequence of interactions between the environment, a person's *resources* (human capital, skills, abilities), and a person's *affordances* (needs, desires, choices). Stokes, Mumford, and Owens (1989) demonstrated that "prototypes" of individuals could be empirically identified that "create their own organized subenvironments consisting of various activities and experiences with which a self-propagating developmental trajectory...crystallizes, or becomes predictable" (p. 512). Mael (1991) elaborated the ecology model, using social identity theory to explain how biodata items tap situations in which "a person associates with a...psychological group," taking on "(to

varying degrees) the syndrome of aspirations, preferences, values, and self-perceptions that are endemic to group members" (p. 768).

These models are valuable conceptualizations of the construct domain underlying life history items. However, they fail to provide strong guidance for operationalization—that is, explicit direction in how to design paper-and-pencil life history inventories to predict specific criteria. Consequently, characterizations of biodata research as atheoretical empiricism will remain until the loop is closed—that is, until support is found for specific linkages between theory, item content, and criterion performance measures (see Dunnette, 1962; Guilford, 1959; Henry, 1966).[1]

At first glance, rational keying procedures for biodata instruments might provide such a linkage (see Hough & Paullin's chapter in this volume). These procedures involve generation of specific a priori hypotheses about how item responses should be related to a criterion based on subject matter experts or investigators' subjective judgments (see Mitchell & Klimoski, 1982). Unfortunately, none of the applications of rational scoring techniques were derived from nomological networks of hypothesized relationships among constructs (two of which include biographical information and job performance). In other words, rational scoring techniques are not excessively burdened with elaborate theory. This criticism should not be considered too harsh, as the field is generally plagued with a lack of accepted theories that explicitly link the content domains of job and person (Bobko & Russell, 1991; Burke & Pearlman, 1988; Dunnette, 1966).

Yet somehow investigators have been able to generate instruments that consistently yield criterion-related validities among the highest reported in the literature (Reilly & Warech, 1990). Further, Owens and his colleagues have demonstrated that subjects' biodata item responses exhibit coherent factor structures and relationships over time (Mumford & Owens, 1982, 1984; Mumford, Wesley, & Shaffer, 1987; Neiner & Owens, 1982; Owens & Schoenfeldt, 1979; Shaffer, Saunders, & Owens, 1986; Stokes, Mumford, & Owens, 1989). These results suggest that systematically scored biodata items capture consistent aspects of antecedent developmental processes that causally influence subsequent job performance. Simultaneous exploration of construct domains underlying criteria and biodata items may provide the answer to Campbell's (1990) call for a theory of "experience."

This growing body of biodata research caused Fleishman (1988) to recently label biographical information as one of the new frontiers in personnel selection. Numerous authors over the years have called for more theory-based investigations of biodata measurement technologies (Bass, 1990; Dunnette, 1966; Henry, 1966; Owens, 1976; Reilly & Chao, 1982; Toops, 1948, 1959). Meta-analytic results suggest that limited resources can be directed away from efforts to establish criterion-related validity at

every job site. Now more than ever, these research resources can be refocused toward more theory-based efforts that *simultaneously* target prediction and explanation. One of the most promising points of departure into this "new frontier" is at the level of biodata item development.

APPROACHES TO DEVELOPING BIODATA ITEM CONTENT

In this section, specific item generation procedures are described in the context of three theories of individual differences. Item generation procedures outlined by Mumford and Owens (1987) are reviewed, and then two specific techniques are applied to theories of personality, vocational choice, and leadership.

Procedure Behind Item Generation

Mumford and Owens (1987) described six methods or sources of biodata item development: (a) the human development literature, (b) life history interviews with incumbents, (c) typical factor loadings of biodata items, (d) known life history correlates of various job specifications, (e) biodata items with known predictive validities, and (f) items generated from the investigators' general psychological knowledge. The latter three sources of biodata items are derived from existing item pools and/or rely solely on investigators' imagination and subjective judgments. Though useful for developing criterion-valid selection instruments, tapping existing biodata item pools without some theoretical rationale does not enhance our understanding of why observed criterion-related validities exist.[2] Though these three procedures will undoubtedly be used in the iterative nature of theory testing, item development, and theory development, it is unlikely that they will provide useful *initial* points of departure for theory development and criterion prediction.

In contrast, the first three sources provide rich sources of information for biodata development. The first three sources reflect two thrusts: theory (the human development literature) and procedure (interviews and factor interpretations).[3] This section attempts to link existing *theories* of individual differences with *procedures* for biodata item development. No attempt was made at being comprehensive or even representative in choosing theories related to job performance. Consequently, I sampled from literatures in personality theory, vocational interests and job choice, and human development/leadership. Theories of personality and vocational interest were chosen as two points of departure because they are convenient: They both

are noncognitive with histories of scale development similar to those of biodata (Barge, 1988). The leadership literature (and the emerging human development literature focusing on leader development) was chosen because theoretical approaches to leadership are arguably as close to a theory integrating construct domains found in the job and individual as we have in the field of industrial and organizational psychology. Further, recent work by Kuhnert (Kuhnert & Lewis, 1987; Kuhnert & Russell, 1990; Russell & Kuhnert, 1992) provided a model of how leaders develop *and* guidelines on how to develop biodata items that reflect critical stages of that development.

Following discussion of each theory or construct domain, specific *procedural* suggestions and examples of how to generate biodata items are provided. Procedures including post hoc interpretation of item factor structures and life history interviews/essays are described as alternate techniques for actual item generation.

Alternatively, equally viable points of departure (theory-based and nontheory-based) could have been chosen from many other arenas. For example, the job constructs of involvement with data, people, or things could be used as a basis for item development procedures. One could also target specific knowledge, skills, and abilities (KSAs) required of a job, working backward to identify the prior life events that influence their acquisition in much the same way that the typical structured interview is conducted (see Russell, Mattson, Devlin, & Atwater, 1990). However, the sheer volume of approaches to jobs and people precludes an exhaustive survey of alternate theories or models that could serve as points of departure for biodata item generation. The topic areas chosen serve as good examples for outlining item generation procedures.

Theory-Based Methods for Developing Item Content

Personality Theory

Most investigators in personnel selection abandoned personality tests long ago, based on conclusions drawn by Guion and Gottier (1965) and the discouraging results of Ghiselli's (1966, 1973) surveys of criterion-related validity. Only recently has interest been revived among applied investigators. For example, efforts by Pulakos, Borman, and Hough (1988) indicate that personality scales chosen on the basis of careful job analysis procedures can yield meaningful incremental criterion-related validities.

Briefly by way of review, Allport (1937) and Murray (1938) conceived of personality in terms of traits manifesting themselves as consistencies in behaviors across a variety of situations. Without reviewing the trials and tribulations that have characterized personality theory over the last 50

years, five personality characteristics that consistently emerge across studies are now viewed as capturing most of the variance in existing measurement instruments.[4] These Big Five characteristics include Extroversion/ Surgency, Agreeableness, Conscientiousness, Emotional Stability, and Culture (see Digman, 1990; Digman & Inouye, 1986). A recent meta-analysis of 49 criterion-related validity studies using the Big Five personality characteristics resulted in average validities ranging from .22 for Conscientiousness to .08 for Agreeableness and Emotional Stability, again indicating that it may be premature to abandon these measures (Barrick & Mount, 1991).

Regardless of their criterion-related validity in selection contexts, the Big Five have substantial evidence of construct validity (Digman & Inouye, 1986). This evidence takes the form of interpretable factor structures, consistency in factor structures over time, and convergent and discriminant validity (see Digman, 1990; Digman & Inouye, 1986). Given that these Big Five behavioral consistencies exist and are related to job performance, how can we generate examples of antecedent life events that (a) are related to differential rates of developing Big Five personality characteristics (e.g., demonstrate convergent and discriminant validity among the Big Five) and (b) demonstrate criterion-related validity?

As noted earlier, the techniques described here will involve factor interpretation and life history interviews/essays. Mumford and Owens described factor interpretation in terms of biodata item factor structures. However, in this application we are faced with factors of personality scale items. It should be feasible to use the Big Five factors (or, for that matter, any other post hoc interpretations of empirically derived factors) to develop biodata items. For example, at least two of the Big Five seem to be capturing some notion of affect (e.g., Agreeableness and Emotional Stability). Using simple notions of frequency of exposure to an experience as well as inputs, process, and outcomes of that experience, it is not difficult for an investigator to generate examples of past situations that might influence these two factors.

Specifically, consider the following items derived for purposes of selecting a retail store manager:

- How often have you been very unhappy with some product or service you purchased for your home? (exposure)
- How often has a clerk or salesperson said something that really irritated you? (input)
- How often have you been very nervous or tense in helping a customer who had a complaint? (process)
- How often have you resolved to try to remain calm after getting upset about something a co-worker did? (outcome)

The position of retail store manager was chosen just to provide a point of reference for the applicant pool (e.g., those out of school who have had some meaningful work/life experiences prior to being considered for the position). Exposure, inputs, process, and outcomes were chosen as an initial taxonomy of ways people might differ in their prior experiences that can be expanded or contracted as item generation proceeds. The taxonomy is taken, with minor changes, from Campbell, Dunnette, Lawler, and Weick's (1970) person-process-product model of managerial performance (see Russell & Domm, 1990, for an initial use of this taxonomy to generate biodata items). Regardless, each item has a specific hypothesized construct that it should be related to (i.e., the notions of happiness, irritation, and resolution to remain calm found in these items would be expected to capture aspects of Agreeableness and/or Emotional Stability). Further, each item has constructs that it should not be related to (the other three members of the Big Five) *and* a content domain (customers and co-workers) that should overlap with aspects of the criterion job performance domain (working with others in a retail sales work environment).

This kind of investigator-dependent item generation can be engaging to the investigators (try thinking of some items related to personal integrity or any deeply held value about "right" and "wrong"—we will come back to this in the section on leadership). However, there is always the risk of contamination or deficiency due to limits of the investigators' imaginations. An alternate procedure that I prefer involves the use of life history interviews or essays with incumbents, hence, shifting the role of subject matter expert (SME) to individuals who are closer to having actually experienced any biodata constructs of interest.[5] Russell, Mattson, Devlin, and Atwater (1990), Russell and Domm (1990), and Siegel (1956) demonstrated that criterion-valid items can be generated from essays written by incumbents about prior life experiences. Russell (1990) demonstrated how criterion-valid life history information can be systematically obtained through tape-recorded life history interviews. Essay and interview questions can be structured to target facets of life episodes that incumbents feel influenced their capacity to perform their jobs. These questions would address environmental circumstances (e.g., task requirements, availability of resources, presence of obstacles, assistance received from others), cognitions (perceptual processes, information gathered, ways information was combined), affect (attitudes, beliefs, values, valences present when the episode started and/or after the episode was complete), behaviors engaged in, task outcomes, and what was learned from the life episode (cognition, affect, and behavior change). Russell et al. (1990) described specific follow-up questions focusing on aspects of the target

job of U.S. naval officer, whereas Russell (1990) used 42 pages of structured life history interview questions targeting job choice, job environment, major accomplishments, major disappointments, obstacles encountered/overcome, specific behaviors engaged in, what was learned from each job tour, and any affect associated with job experiences.

In the case of the Big Five Extroversion/Surgency scale, we could ask incumbent retail store managers who had been recently hired (i.e., those most similar to the applicant pool) to describe prior life experiences (work-related or nonwork-related) in which they had been required to be particularly vocal, secretive, cautious, adventurous, or sociable (aspects of the Extroversion/Surgency dimension). Again, each essay or interview could be structured to prompt respondents for descriptions of the circumstances in which they found themselves, what they thought and felt at the time, what they did, what outcome occurred, and what they feel they learned from the situation. Thus, a pool of life experiences can be generated that *are not* dependent on the imaginations of the investigators.

Regardless of the technique used, it would appear that the Big Five personality variables are capturing behavioral consistencies that demonstrate criterion-related validity *and* construct validity (Barrick & Mount, 1991; Digman, 1990). The techniques just described provide a means of developing biodata items that might be expected to augment both construct- and criterion-related validities. I would not be surprised if future investigators find that a subset of past behavioral consistencies common to work-related roles (as well as their antecedents and consequences) exhibits much higher criterion-related validity than that reported by Barrick and Mount (1991), yet is readily interpretable in terms of the Big Five constructs.

Vocational Interest and Job Choice

Holland's (1973) model of vocational choice is a direct application of personality theory. Briefly, Holland believed that an individual's similarity to six personality types—Realistic, Investigative, Artistic, Social, Enterprising, and Conventional—could be used to place the individual into vocational groups. Similar to Mumford and Stokes' (1991) ecology model, Holland's (1973) view is that individuals' choices of a vocational type is a function of their environments, personal abilities, and desires.

Osipow (1973) criticized Holland's theory for not indicating how these personality types developed, though some research has addressed this issue. A large number of studies in the early 1960s examined how biodata items were related to the career paths of engineers, lawyers, physicians, and other professionals (see Albright & Glennon, 1961; Chaney & Owens, 1964; Kuhlberg & Owens, 1960). Not surprisingly, engineers and lawyers tended to have different histories of interpersonal success and differential

performance in quantitative versus language courses (see Kuhlberg & Owens, 1960).

More recently, Eberhardt and Muchinsky (1982, 1984), in a large-scale survey, found that prior life experiences captured through a biodata instrument can accurately predict vocational type. Holland's use of personality characteristics to identify vocational types lends itself to the biodata item generation procedures described in the previous section. Given meta-analytic results suggesting that biodata instruments are among the most accurate predictors of voluntary turnover, it would appear that life history items are *also* stable predictors of the motivational states reflected in vocational and job choices.

Placing the motivational process of vocational and job choice into an expectancy theory framework provides a different point of departure for the identification of life history events. Items might be developed that reflect events influencing individuals' expectancies, instrumentalities, and valences in some prior work or goal-oriented activity. Although Kuhlberg and Owens (1960) report differences in prior life experiences between engineers and lawyers, it would be of theoretical and practical interest to identify which facets of prior developmental processes are actually related to the way people choose jobs or career paths.

For example, it would be interesting to know which situations—home, school, part-time job—have the greatest impact on work-related expectancies, instrumentalities, and valences (EIV) in high school and college students. What kind of role models at home, school, or part-time employment have the greatest impact on EIV? Is it the mere presence of these role models, or is some opportunity for a particular *type* of role model interaction necessary? Again, life history essays targeting role models and role model interaction could be used to obtain an initial set of biodata items. Relationships between items reflecting specific role model characteristics and environmental circumstances with work-related EIV would be of great value for theory development, vocational guidance, and personnel selection.

Equally interesting to applied researchers *and* employers is the exact nature of early developmental experiences that contribute to later career performance. For example, companies like General Electric and Westinghouse would be interested in knowing which approach to performance management influences EIV that cause engineers to continue as practicing engineers rather than opting for career tracks in technical sales or management.

In the context of vocational decisions to leave a job or career path, Russell and Van Sell (1986) demonstrated that within-subjects policy-capturing research designs are much more likely to accurately forecast voluntary turnover decisions than the currently popular between-subjects

longitudinal panel surveys (e.g., Williams & Hazer, 1986). Russell and Van Sell developed regression models of how each employee weighs and combines various facets of his or her job in arriving at a decision to quit. Further, using a cluster analysis technique, they were able to identify groups of incumbents who were similar in the relative weights they assigned to different aspects of the job. For purposes of selection/classification (requiring prediction) and career guidance (requiring explanation), it would be of interest to identify different biographical experiences that influence these weights.

For example, the following items might be expected to be related to how individuals weigh scheduling flexibility and pay fairness in their decisions to quit a job:

- How often did you miss your parents when you were growing up because they were away at work? (H_0: affects valence of flexible work schedules.)
- How often while you were in school did you work part-time jobs but still didn't have as much money as you needed? (H_0: affects development of a "need" norm and may affect the valence perceptions of subsequent pay levels.)
- How often have you continued to work at something even though you weren't getting the reward you originally thought you were going to get out of it? (H_0: affects development of capacity to identify other valent outcomes in a work situation.)

Again, my preference would be to use a combination of life history essays and interviews targeted at prior life events that affect motivation to pursue a particular line of work or to continue on a job even after conditions of work have changed. The following essay questions, given to first-year college students, might yield responses that can be used to develop items demonstrating construct- and criterion-related validity with subsequent measures of career choice/job choice/turnover:

- Please describe some job or task you have worked on (related to work, school, or any other situation you may have encountered) in which the situation changed yet you continued to participate. Please be sure to describe what things changed, how you *initially* felt about the changes, how you *ended up* feeling about the changes, why you thought you could deal with the changes, and how the changes influenced what you were getting out of the situation.
- Please describe some situation in which you had to decide between different types of projects or activities in which to take part. Examples would include choosing among elective courses in school, choosing

among different summer jobs, or choosing among different colleges to attend. These choices can be very difficult for some people. How did you make yours? How did you know what you wanted? How did you know you could do it? How did you know what to expect from each alternative? What did you learn that will prompt you to do things differently the next time you have to make a choice like this?

By decomposing the events into sequences of exposures to different environments and responses to the environment (e.g., cognitive, affective, and behavioral components of attitudinal response as described by Kiesler, Collins, & Miller, 1969, and Kretch & Crutchfield, 1948), we should be able to generate biodata items that are accompanied by specific hypothesized relationships with subsequent levels of motivation that are directly reflected in career and job choices.

Leadership and Human Development Theory

Kuhnert and his colleagues recently used a theory of adult development to describe how individuals acquire skills as transactional and transformation leaders (Kuhnert & Lewis, 1987; Kuhnert & Russell, 1990; Russell & Kuhnert, 1992). Using Kegan's (1982) model of adult development, Kuhnert and Lewis (1987) distinguished between the concepts of *object* and *subject* at different stages of development. The process or structure through which an individual makes sense of his or her experiences in the world is called the *subject* . The metaphor most commonly used in referring to the subject is that of a lens. In contrast, the content of the experiences (what is being viewed through the lens) is called the *object.* Kegan (1982) hypothesized that as humans mature, the way in which they make sense of things (subject) and the things they pay attention to (object) change. The six stages of development described by Kegan are characterized by the subject of earlier stages becoming the object of later stages.

For instance, at Stage 3, Kegan describes the subject in terms of interpersonal mutuality—the rules of interaction or exchange (e.g., equity, equality) that individuals use to view the world. The object is the individuals' knowledge of their own and others' needs, wishes, and interests. Hence, Kuhnert and Lewis (1987) argued that the Stage 3 adult has the foundation needed to be a transactional leader—one who influences followers by managing contingencies.

At Stage 4, Kegan (1982) described individuals who have developed to the point where they can stand back and view the rules of the exchange—what was subject in Stage 3 becomes object in Stage 4. The new subject consists of deep-seated values or ideology held by individuals through which they view the exchange. Hence, whereas in Stage 3 the content of

the exchange was viewed through a lens consisting of the rules of exchange, in Stage 4 deep-seated values make up the lens used to view the rules of exchange. A Stage 3 individual would have difficulty violating a rule of exchange in some interaction with a subordinate, whereas a Stage 4 individual would be able to violate the rule of exchange *if that violation was congruent with some deep-seated value.* For example, Stage 4 leaders would routinely violate some trust with a subordinate or take action that was detrimental to their own good because of some overarching, higher-level value (e.g., doing what is "right" versus doing what is in the leader's or others' best interests).

Kuhnert and Lewis (1987) used these stages of development as a framework to describe how people acquire transactional and transformational leadership skills. Transformational leaders influence others by inducing a relatively permanent shift in their values, beliefs, and goals—that is, their beliefs about what is right (Stage 4). The transactional leader influences others through the careful management of inducements and contributions to meet both the organization's and employees' needs (Stage 3).

Kuhnert and Russell (1990) argued that how people make sense out of their past life experiences should reflect their stage of development. A logical means of operationalizing what is subject and what is object—a person's stage of development—is to use items that capture aspects of that person's life history. They use an existing criterion-valid biodata item taken from Russell et al. (1990) as an example:

> On a group project, many times a person is not pulling his or her own weight. Sometimes you have to discuss this with the person. How often did these discussions "work out" and resolve the problem?

Kuhnert and Russell (1990) argued that different ways of viewing and responding to this situation should discriminate between Stage 3 and 4 leaders. Two subsequent items provided by Kuhnert and Russell that describe the same situation but in which the item ends with either "I pointed out it wasn't fair to the other individuals in the group" or "I explained how the individual was letting him or herself down" should yield different response patterns from Stage 3 and 4 leaders. Stage 3 leaders should indicate they often engage in the former action and seldom in the latter action. Stage 4 leaders should indicate they engage in each action with about equal, and high, frequency.

These constitute examples of biodata items accompanied by explicit hypotheses derived from established psychological theories of adult development and leadership. Again, post hoc interpretation of factor structures, as well as life history essays, could be used to ensure a rich pool of items reflecting critical developmental episodes that might signal points of

change from one stage to the next. For example, Russell et al. (1990) and Russell and Domm (1990) interpreted the dominant biodata factor in their item generation efforts as capturing life problems or negative life experiences. Interestingly, recent ethnographic work on key events in executives' lives by Bobko, McCoy, and McCauley (1988) and Lindsey, Holmes, and McCall (1987) indicates that having worked under a very stressful, obnoxious boss (an unpleasant life experience) is commonly cited as a key developmental period by many upper-level managers. Future item development efforts might ask incumbents for life history essays about aspects of negative life events.

ITEM CHARACTERISTICS

In this final section I shall discuss issues related to item characteristics that are independent of item content. Item characteristics are interpreted broadly to include, for example, characteristics of the accompanying instructions, choice of response formats, and item tone (negative versus positive; see Asher, 1972, for an overview of different item characteristics). I shall discuss the relationship of these features to random and systematic response error and item validity, though few definitive studies have addressed these issues.

Impact on Reliability

Numerous reasons for the presence of error in biodata item responses have been suggested. Van Rijn (1980) listed faking, errors in memory, carelessness, and response bias as potential sources. Response biases might include selective memory of certain types of life experiences and maturation effects that cause changes in how prior perceptions, cognitions, attitudes, values, or beliefs are recalled at a later point in time. However, as we shall see, many of these biases may indeed represent error variance resulting from meaningful life experiences that occur at a later date, as revealed in Kuhnert and Lewis' (1987) discussion of changes in subject and object during leader development.

The heterogeneity of biodata items usually dictates that split-half reliabilities and other estimates of internal consistency reliability generally will be low (see Owens, 1976; Siegel, 1956). Additionally, if England's (1961) vertical percent difference method of item scoring is employed (which, as Devlin, Abrahams, & Edwards, 1992, demonstrated, yields the highest cross-validities compared to five competing methods), high test–retest correlations could conceal a shift in mean response. Such a shift

would have a drastic effect on resultant biodata scores obtained using England's procedure.

Owens, Glennon, and Albright (1962) avoided these concerns by examining 43 subjects' response consistency to 200 items. An interval of approximately two months lapsed between administrations. Owens et al. developed four rules or principles to describe why some items received consistent (or inconsistent) responses. These rules referred to item brevity, graduated or continuous response scales, presence of an "escape" response, and the use of neutral or pleasant undertone to the item stem. To my knowledge, none of these post hoc interpretations of why some items yielded consistent responses has been evaluated for its relationship to item reliability or validity.

Shaffer, Saunders, and Owens (1986) replicated and extended this effort in a test–retest design over a five-year period combined with an independent measure of life experiences. They examined both test–retest reliabilities and *t* tests of differences in mean value for factor scores and item responses. Shaffer et al. (1986) reported consistent evidence of high test–retest reliability for both factors and items. However, approximately one-third of the factors demonstrated significant shifts in mean response. Slightly over 30 percent of the items demonstrated a significant shift in mean response over the five-year period, while over 26 percent of the items were significantly different from an independent assessment of the same life history experiences. Shaffer et al. also presented evidence suggesting that objective to moderately subjective items (as opposed to highly subjective items) are not likely to demonstrate mean shifts.

Unfortunately, in designs of this nature, it is very difficult to separate random error (e.g., careless responding, failure of memory, etc.) from Campbell and Stanley's (1963) description of history and maturation effects. That is, lack of test–retest reliability and/or shifts in mean response for biodata items or factors may result from specific events that occur between administrations or from changes in the respondent. History and maturation effects are exactly what is hypothesized by Mumford and Stokes' (1991) ecology model as the causal process underlying the predictive power of biodata items.

Hence, it may be that items exhibiting the most change in how prior life experiences are recalled over time are those with the highest criterion-related validity, simply because these items are more reflective of critical developmental episodes. It is not surprising that studies of accuracy in biodata item responses using test–retest procedures and/or independent verification yield mixed results (Mumford & Owens, 1987). Simple shrinkage in criterion-related validity upon cross-validation should measure the amount of random error in biodata item responses. I am unaware of any studies examining the effect of manipulating item characteristics on

shrinkage in cross-validities. Items influenced by history and maturation effects that truly contribute to validity but detract from traditional measures of reliability (e.g., Shaffer et al.'s test–retest reliabilities and *t* tests) should survive cross-validation.

Finally, a discussion of the impact of item characteristics on reliability would not be complete without noting widespread concerns about one source of nonrandom error, that is, faking (see Fleishman, 1988). Numerous authors have reported evidence of faking (Cascio, 1975; Colquitt & Becker, 1989; Keating, Paterson, & Stone, 1950; Mosel & Cozan, 1952), while others have failed to find evidence (Goldstein, 1971; Hogan & Stokes, 1989; Trent, Atwater, & Abrahams, 1986; Weiss & Dawis, 1960). Indeed, Trent (1987) found that the ability of respondents to detect and respond to the keyed option contributed to criterion-related validity regardless of what their true response might have been. Apparently, in some applicant pools the cognitive capacity to detect and fake the desired response correctly forecasts performance.

In the face of these mixed results, some authors have suggested that the only way to eliminate faking is to rely on verifiable biodata items. However, Mumford and Owens (1987) speculated that choice of keying procedure (item keying versus option keying) may be causing the mixed results. A recent lab study by Kluger et al. (1991) tested this speculation, finding that option keying caused biodata scores not to be inflated due to bias caused by knowledge of the target position and general social desirability, though such biases did contribute to random error variation. Item keying, however, did result in inflated biodata scores. Kluger et al. (1991) suggested that to reduce inflation of random error, future researchers should examine the effect on response bias of instruction sets explicitly warning respondents that attempts to fake will not be fruitful.

Impact on Validity

Few investigators have evaluated how item characteristics affect criterion-related validity. Recently, Barge (1987, 1988) reported the results of a study examining how three biodata item characteristics are related to item validity independent of item content. Barge found that heterogeneity, behavioral discreteness, and behavioral consistency of biodata items could be reliably rated and, using job performance and training criteria, found that these properties were related to criterion-related validity.

Item heterogeneity refers to the distinction between items that capture multiple characteristics or environmental events (e.g., school performance) versus items that narrowly reflect a single characteristic. *Behavioral discreteness* refers to the distinction between items that address "a single, perhaps

verifiable, behavior rather than a more abstract or summary characteristic" (Barge, 1987, pp. 3–4). Finally, *behavioral consistency* refers to the degree of congruency between the content domain of the biodata item and the content domain of the target job—that is, the degree to which the item is a *sample* as opposed to a *sign* (Wernimont & Campbell, 1968).

While Barge (1987, 1988) found real differences in the criterion-related validity of items varying on these dimensions, it remains to be seen whether these differences are replicated with different sets of items. Specifically, Barge used 103 items out of the 118 items in the short form of Owens' *Biographical Questionnaire* (BQ; Owens & Schoenfeldt, 1979). This instrument was derived from an initial set of 2,000 items generated to reflect 52 pages of item categories (e.g., dependency, aggression) that the team of investigators viewed as capturing a broad array of inputs and behaviors. Do item heterogeneity, behavioral discreteness, and behavioral consistency influence construct- and criterion-related validity for items reflecting life history construct domains other than those found in the Owens' *Biographical Questionnaire* (e.g., those relevant for the selection of middle managers)? Future research will need to examine simultaneously both differences in theoretical rationale for item content and variability in item characteristics. It may be that a number of items with characteristics exactly opposite from those reported by Barge (e.g., homogeneous, nondiscrete sign items) are needed to accurately measure certain critical constructs or life events that are not reflected in the short form of the BQ.

CONCLUSION

The focus of this chapter has been on the generation of biodata items with content that can be traced to some theory or model. The goal of such efforts is to close the gap between theory, item content, and criterion performance measures. Theories focusing on person characteristics (personality theory), motivation (vocational choice), and the development of individuals into leaders were discussed. Any one of a number of alternative approaches could have been chosen, and the reader should consider it a challenge to identify causal, developmental life experiences that will confirm (or fail to confirm) his or her favorite theory or model.

Life history essays/interviews and post hoc factor interpretations have been featured as means of item development—obviously, many others are possible. I have been fascinated by autobiographies that describe major life-influencing events. For example, Armand Hammer (1987) attributed many of his early career decisions to the circumstances surrounding his father's

illness. Dr. An Wang (1986) paused to reflect on basic values of right and wrong while negotiating with IBM on the sale of rights to his memory cores in the early 1950s.

Owens (1976) commented on the possibility of abstracting items from biographies, but expressed concern that it would not be an efficient method of item generation. Fortunately, microcomputer-based text search software has been developed that, like data base searches in library archives, could be used to target passages of text dealing with specific themes and issues (see Gephardt & Wolfe, 1989). Imagine the biodata sampling possibilities such a tool would provide when combined with the computerized biography section of the Library of Congress!

Research indicates that personality, vocational choice, and leadership models, among others, provide fairly accurate representations of nonrandom individual differences. Unfortunately, many empirical examinations of these theories and others conclude with a statement that "the results are consistent with the theory," even though no criterion of organizational interest was predicted. In contrast, biodata inventories consistently demonstrate criterion-related validities as high as any competing predictor. We need to close the gap between what has been characterized as an atheoretical measurement technology and the many substantive theories of industrial and organizational psychology. Toward this end, methods of theory-based biodata item generation comprise a major research opportunity.

A portion of this chapter was written while I held a visiting appointment at the Krannert Graduate School of Management, Purdue University. I would like to thank Philip Bobko and Marvin Dunnette for their valuable comments and suggestions.

NOTES

1. This criticism is common to many "noncognitive" domains and is not unique to biodata research. For example, Eberhardt and Muchinsky (1982) made a similar observation regarding Holland's (1973) use of personality types to explain underlying vocational interest—that is, that viewing personality at any point in time as a function of genetic and current/past environmental factors is too general to be of much use.

2. Nickels and Mumford (1989) recently completed a study examining the construct- and criterion-related validity obtained in using a priori theory-based guidelines to select biodata items from an existing pool. While this is a perfectly viable approach, the focus of this chapter is on biodata item development, not biodata item selection.

3. Though I would argue that the sixth source of items, investigators' general psychological knowledge, permeates the other five sources, especially in the interpretation of factor loadings.

4. See the chapter by Mary Tenopyr in this volume for a more extensive treatment of the roots of the Big Five.

5. I faced this problem when trying to decide how to develop items capturing biographical information for purposes of selecting midshipmen for the U.S. Naval Academy (Russell, 1986). As a member of the thirty-something generation, I was *very* uncomfortable with the idea of trying to generate item content reflective of life history events for teenagers today.

REFERENCES

Albright, L. E., & Glennon, J. R. (1961). Personal history correlated of physical scientists' career aspirations. *Journal of Applied Psychology, 45,* 281–284.

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt, Rinehart and Winston.

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

Bass, B. M. (1990). *Bass and Stodgill handbook of leadership* (3d ed.). New York: Free Press.

Barge, B. N. (1987, August). *Characteristics of biodata items and their relationships to validity.* Paper presented at the annual meeting of the American Psychological Association, New York.

Barge, B. N. (1988). *Characteristics of biodata items and their relationship to validity.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Barrick, M. R., & Mount, M. K. (1991). The Big 5 personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Bobko, P., McCoy, C., & McCauley, C. (1988). Towards a taxonomy of executive lessons: A similarity analysis of executives' self-reports. *International Journal of Management, 5,* 375–384.

Bobko, P., & Russell, C. J. (1991). A review of the role of taxonomies in human resource management. *Human Resource Management Review, 1,* 293–316.

Burke, M. J., & Pearlman, K. (1988). Recruiting, selecting, and matching people with jobs. In J. P. Campbell & R. J. Campbell (Eds.), *Productivity in organizations* (pp. 97–142). San Francisco: Jossey-Bass.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs.* Chicago: Rand McNally.

Campbell, J. L. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2d ed., vol. 1, pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J., Dunnette, M. D., Lawler, E. E., & Weick, K. (1970). *Managerial behavior and performance effectiveness.* New York: McGraw-Hill.

Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology, 60,* 576–580.

Chaney, F. B., & Owens, W. C. (1964). Life history antecedents of sales, research, and general engineering interests. *Journal of Applied Psychology, 48,* 101–105.

Colquitt, A. L., & Becker, T. E. (1989, April). *Faking of a biodata form in use: A field study.* Paper presented at the meeting of the Society of Industrial and Organizational Psychology, Boston.

Devlin, S. E., Abrahams, N. M., & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. *Military Psychology, 4,* 119–136.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50,* 116–123.

Dunnette, M. D. (1962). Personnel management. *Annual Review of Psychology, 13,* 285–314.

Dunnette, M. D. (1966). *Personnel selection and placement.* Belmont, CA: Wadsworth.

Eberhardt, B. J., & Muchinsky, P. M. (1982). 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: Brown.

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

Gephardt, R.P., & Wolfe, R.A. (1989, August). Qualitative data analysis: Three microcomputer-supported approaches. In *Academy of Management Best Papers Proceedings 1989.* Washington, DC: Academy of Management.

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

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology, 26,* 461–477.

Goldstein, I. L. (1971). The application blank: How honest are the responses? *Journal of Applied Psychology, 55,* 491–492.

Guilford, J. P. (1959). *Personality.* New York: McGraw-Hill.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135–164.

Hammer, A. (1987). *Hammer.* New York: Putnam.

Henry, E. (1966). Conference on the use of biographical data in psychology. *American Psychologist, 21,* 247–249.

Holland, J. L. (1973). *Making vocational choices: A theory of careers.* Englewood Cliffs, NJ: Prentice-Hall.

Hogan, J. B., & Stokes, G. S. (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 annual meeting of the Society for Industrial and Organizational Psychology, Boston.

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

Keating, E., Paterson, D. G., & Stone, C. H. (1950). Validity of work histories obtained by interview. *Journal of Applied Psychology, 34,* 1–5.

Kegan, R. (1982). *The evolving self: Problem and process in human development.* Cambridge, MA: Harvard University Press.

Kiesler, C. A., Collins, B. E., & Miller, N. (1969). *Attitudinal change: A critical analysis of theoretical approaches.* New York: Wiley.

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76,* 889–896.

Kretch, D., & Crutchfield, R. S. (1948). *Theory and problems in social psychology.* New York: McGraw-Hill.

Kuhlberg, G. E., & Owens, W. C. (1960). Some life history antecedents of engineering interests. *Journal of Educational Psychology, 51,* 26–31.

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,* 1–13.

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

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44,* 763–792.

Mitchell, T. W. (1992, April). *A priori biodata to predict teamwork, learning rate, and retention.* Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Montreal.

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

Mosel, J. L., & Cozan, L. W. (1952). The accuracy of application blank work histories. *Journal of Applied Psychology, 36,* 365–369.

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, 11,* 1–31.

Mumford, M. D., & Stokes, G. S. (1991). Developmental determinants of individual action: Theory and practice in the application of background data. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2d ed., vol. 3, pp. 1–78). Palo Alto, CA: Consulting Psychologists Press.

Mumford, M. D., Wesley, S. S., & Shaffer, G. S. (1987). Individuality in a developmental context II: The crystallization of developmental trajectories. *Human Development, 30,* 291–321.

Murray, H. A. (1938). *Explorations in personality.* New York: Oxford University 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.

Nickels, B. J., & Mumford, M. D. (1989). *Exploring the structure of background data measures: Implications for item development and scale validation.* Unpublished manuscript.

Osipow, S. H. (1973). *Theories of career development* (2d ed.). Englewood Cliffs, NJ: Prentice-Hall.

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. Dunette (Ed.), *Handbook of industrial and organizational psychology* (1st ed., pp. 609–644). Chicago: Rand McNally.

Owens, W. A., Glennon, J. R., & Albright, L. E. (1962). Retest consistency and the writing of life history items: A first step. *Journal of Applied Psychology, 46,* 329–331.

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64,* 569–607.

Pulakos, E. D., Borman, W. C., & Hough, L. M. (1988). Test validation for scientific understanding: Two demonstrations of an approach to studying prediction–criterion linkages. *Personnel Psychology, 41,* 703–716.

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

Reilly, R. R., & Warech, M. W. (1990). *The validity and fairness of alternatives to cognitive tests.* Berkeley, CA: Commission on Testing and Public Policy.

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

Russell, C. J. (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: Navy Personnel Research and Development Center.

Russell, C. J. (1990). Selection of 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 method of item generation.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Miami.

Russell, C. J., & Kuhnert, K. W. (1992). Integrating skill acquisition and perspective taking capacity in the development of leaders. *Leadership Quarterly, 3,* 109–135.

Russell, C. J., Mattson, J., Devlin, S. E., & Atwater, D. (1990). Predictive validity of biodata items generated from retrospective life experience essays. *Journal of Applied Psychology, 75,* 511–523.

Russell, C. J., & Van Sell, M. (1986, August). *An examination of the processes underlying models of job choice and turnover* (winner of the Ghiselli Award for Research Design). Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37,* 407–422.

Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical information: Long-term retest and observer ratings. *Personnel Psychology, 39,* 791–809.

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

Stokes, G. S., Mumford, M. D., & Owens, W. A. (1989). Life history prototypes in the study of human individuality. *Journal of Personality, 57,* 509–545.

Toops, H. A. (1948). The use of addends in experimental control, social census, and managerial research. *Psychological Bulletin, 45,* 41–74.

Toops, H. A. (1959). A research utopia in industrial psychology. *Personnel Psychology, 12,* 189–225.

Trent, T. (1987, August). *Armed forces adaptability screening: The problem of item response distortion.* Paper presented at the annual meeting of the American Psychological Association, New York.

Trent, T. T., Atwater, D. C., & Abrahams, N. M. (1986). *Biographical screening of military applicants: Experimental assessment of item response distortion.* Paper presented to the Tenth Psychology in the DOD Symposium, Denver.

van Rijn, P. (1980). *Biographical information and scored application blanks in personnel selection.* Washington, DC: U.S. Office of Personnel Management, Personnel Research and Development Center, Alternatives Task Force.

Wang, A. (1986). *Lessons: An autobiography.* Reading, MA: Addison-Wesley.

Weiss, D. J., & Dawis, R. V. (1960). An objective validation of factual interview data. *Journal of Applied Psychology, 44,* 381–385.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52,* 372–376.

Williams, L. J., & Hazer, J. T. (1986). Antecedents and consequences of satisfaction and commitment in turnover models: A reanalysis using latent variable structural equation methods. *Journal of Applied Psychology, 71,* 219–231.