JOEL P. WIESEN, Ph.D., DIRECTOR
APPLIED PERSONNEL RESEARCH
6 CAULFIELD CIRCLE
NEWTON CENTRE, MA 02159

Personnel Assessment Monographs

■

Biodata: Potentials and Challenges in Public Sector Employee Selection

■

Volume 2, Number 4
April 1992

ipmaac

A publication of the Assessment Council of the International Personnel Management Association.

Biodata: Potentials and Challenges in Public Sector Employee Selection

Paul van Rijn

A publication of the Assessment Council of the International Personnel Management Association.

BIODATA: POTENTIALS AND CHALLENGES IN PUBLIC SECTOR EMPLOYEE SELECTION

Paul van Rijn, Ph.D.

Personnel Assessment Monographs is published by the Assessment Council of the International Personnel Management Association (IPMAAC), 1617 Duke Street, Alexandria, Virginia 22314. A copy is provided without charge to members of the Assessment Council. Others may obtain copies by writing the Director of Assessment Services at the above address. The monographs are intended to provide the most current information available on topics of broad interest in the field of applied personnel assessment. Manuscripts and proposals are solicited by the editor. Most appropriate are reviews of research on assessment methods and innovative applications and solutions to personnel assessment problems in employee selection, performance appraisal, program evaluation, management analysis and related areas which are not generally available elsewhere. Manuscripts are reviewed by the editor and by consulting editors selected for their expertise in the area covered by a particular paper. Papers are selected for publication based on the technical and professional merit of the work, its applicability to current issues, and its utility and interest to the targeted readers. Inquiries, proposals and manuscripts may be sent to the editor, Linsey J. Craig, Palm Beach County, 3111 South Dixie Highway, #120, West Palm Beach, Florida 33405.

Linsey J. Craig, Editor

EXECUTIVE SUMMARY

Recent years have seen an increasing interest in the use of biodata for employee selection. This interest has not escaped the public sector. Some reasons for the increased interest in biodata include: (1) the persistence with which adverse impact continues to accompany many traditional written tests, (2) the desire to expand the scope of predictors beyond the domain of cognitive abilities alone, and (3) promising advances in the understanding and development of biodata questionnaires.

"Biodata" (short for biographical data) refers both to a type of information *about* applicants as well as to a powerful and flexible selection technology. Biodata generally is collected directly from job applicants via highly structured multiple-choice questionnaires and typically includes the kinds of data obtained on job application blanks, in some employment interviews, in personnel files, or in autobiographies.

Unlike most selection procedures, there are no a priori right or wrong answers to the biodata questions. Rather, correct responses on biodata questionnaires are scored according to the extent that they distinguish successful from barely acceptable or unsuccessful employees. Only responses that predict successful job performance are scored and weighted. In effect, biodata is job-related by definition and criterion-related validity is built into its scoring procedure.

Recent research has demonstrated the *potential* benefits to be derived from the use of *properly developed*

CONTENTS (Continued)

Personnel Assessment Monograph

Biodata: Potentials and Challenges in Public Sector Employee Selection

Volume 2, No. 4 *April 1992*

CONTENTS

(Continued)

ACKNOWLEDGEMENTS

No work such as this is born without its debts. It is not possible to identify all the organizations and individuals who helped make this paper a reality. First of all, there was the expressed interest of the International Personnel Management Association (IPMA), followed by the immediate and direct support of the IPMA Assessment Council (IPMAAC), first under the leadership of Dennis Joiner and then Kaye Evleth. An IPMAAC Advisory Committee, consisting of Charles Sproule (Chair) of the Pennsylvania Civil Service Commission, Jay Gandy of the U.S. Office of Personnel Management, and Michael Dollard of the New York State Department of Civil Service provided continuing encouragement and made numerous positive contributions to this paper during every phase.

Special mention, credit, and gratitude must also be extended to Dr. Terry Mitchell for helping shape both the direction and substance of this paper. Additional reviews by Drs. Fred Mael, Hilda Wing, and Charles MacLane added substantially to the quality of this paper. Not to be overlooked are Drs. David Dye, James Sharf, and Frank Schmidt for sharing their ideas and unpublished manuscripts for inclusion in this paper. Finally, the author could not have completed this paper without the understanding and support of his immediate family, friends, and colleagues.

The author accepts full responsibility for any errors, omissions, or oversights. The views expressed do not necessarily reflect those of IPMA, IPMAAC, or the author's primary employer, the U.S. Merit Systems Protection Board.

Paul van Rijn
April 1992

DR. PAUL VAN RIJN is a senior personnel research psychologist for the Office of Policy and Evaluation of the U.S. Merit Systems Protection Board (MSPB) in Washington, DC. While at MSPB, he has conducted nationwide studies of employee turnover in the federal government, culminating in reports on *who* is leaving the civil service and *why*. He is currently preparing a report on the federal government's blue-collar workforce.

Prior to joining MSPB in 1987, Paul was with the U.S. Army Research Institute where he conducted research on organizational productivity, military leadership, and Army civilian personnel issues. He also worked eight years in the Personnel Research and Development Center of the U.S. Civil Service Commission (now the U.S. Office of Personnel Management). During his tenure there, he developed and validated a variety of examinations and conducted early research on the use of self-assessment and biodata in public sector employee selection.

He received his BA from the University of San Francisco and his MA and PhD in experimental psychology from the University of Colorado. He is a member of the American Psychological Association and of the International Personnel Management Association Assessment Council (IPMAAC), and currently serves on the Executive Board of the Personnel Testing Council of Metropolitan Washington, DC.

PREFACE

The IPMA Assessment Council is pleased to present this issue of our *Personnel Assessment Monograph* series in order to provide information and to help guide policy decisions on the development and use of biographical data instruments in personnel selection in the public sector.

In 1990, the Executive Council of the International Personnel Management Association (IPMA) requested the IPMA Assessment Council (IPMAAC) to develop a report on the advantages and disadvantages of using biodata in personnel selection. On the basis of that request, the IPMAAC leadership requested Dr. Paul van Rijn to prepare the report. Dr. van Rijn was selected because he is an active IPMAAC member with keen interest in biodata, having been responsible for the preparation of a valuable reference report on biodata for the United States Office of Personnel Management in 1980. A review committee, chaired by former IPMAAC President Charles F. Sproule and including IPMAAC Board members Michael J. Dollard and Jay A. Gandy, was appointed to provide guidance and assistance to the author, to review and comment on drafts of the report, and make recommendations to the IPMAAC Board of Directors concerning the report.

A summary of the findings of this report was provided to the IPMA Executive Council in March, 1991. The IPMA Assessment Council is now publishing this report in its entirety as a service to its members and the assessment profession.

biodata in public sector selection. These potential benefits include:

1. High test-retest reliability, especially in contrast to personality tests, interest inventories, other noncognitive measures, and typical employment interviews.

2. High criterion-related validity, especially against objective criteria such as employee turnover, which are not as well predicted by many other selection techniques.

3. Validities that are generalizable and relatively stable over time, across different subgroups, occupations, organizations, and situations.

4. Relatively little adverse impact.

5. Good cost-effectiveness in contrast to assessment centers, performance simulation exercises, proficiency tests, and other predictors of behavioral job dimensions, especially when administered to large applicant groups.

6. Easy administration to large groups of job applicants.

7. Relatively non-threatening to most job applicants.

8. Broad and flexible basis for assessment and increased composite validity when used with written tests.

However, would-be users of biodata must proceed with caution. Biodata is not one technology or one type of item. As Mael (1992) points out, any claimed advantages or disadvantages for biodata may only be true for certain types of biodata. Biodata that are factual, objective, and verifiable tend to have excellent psychometric properties and have the potential to deliver many of biodata's benefits. At the other extreme, biographical information about hypothetical situations, feelings, or personality traits of applicants offer fewer, if any, of biodata's potential advantages.

In recent years, even the best biodata procedures have to address some serious challenges on at least two interrelated fronts: (1) acceptability and (2) legality. Today, most applicants would object to many traditional biodata questions (e.g., How many children do you have at home?) on grounds that such questions are not job-relevant, are potentially discriminatory, and are an unwarranted invasion of the applicant's privacy. Increasingly, federal and state laws, regulations, and pre-employment inquiry guidelines are targeting many historically "valid" biodata items as unacceptable. In addition to prohibitions on inquiries related to race, sex, national origin, religious preference, marital status, and political affiliation, questions related to citizenship, finances, personality traits, and family life would be unacceptable in most public sector selection procedures.

There are at least two additional issues that present special challenges for users of biodata: (1) adequate validation designs, and (2) job-relatedness. Typical empirical validation designs for developing the scoring key and conducting the necessary cross-validation generally require large sample sizes (n = 400 to 1,000). To achieve these large sample sizes, jurisdictions with similar occupations should consider combining their resources. However, even the highest quality validation design may be criticized if the job-relatedness of the biodata is not acceptable or understandable to the applicants.

Recent research by Gandy, Outerbridge, Sharf, and Dye (1989) at the U.S. Office of Personnel Management, as well as by others, has demonstrated that many traditional objections to biodata can be overcome by carefully prescreening the questions to be included in the selection procedure. Gandy et al. included only questions that were

verifiable, job-relevant, non-intrusive, under the applicant's control, and non-discriminatory. Although there are variations on this approach, its objective is to build in a rational relationship between biodata and the measure of job success, *in addition to the empirical relationship*. The resulting increase in the acceptability of the biodata questionnaire does not come without a cost, however, since the prescreens may eliminate some of the more reliable and predictive biodata questions, simply because we lack an understanding of why the question predicts job success.

In summary, the biodata technology offers potential benefits as a reliable, valid, and cost-effective selection procedure with little or no adverse impact. It is particularly suitable in situations where there are very large numbers of applicants and relatively few vacancies. However, the use of biodata in the public sector is becoming increasingly complicated—technically, conceptually, and legally. Now more than ever, using biodata requires a high level of expertise across scientific, technical, social, and legal domains. Properly and professionally developed biodata are very much in the interests of employers and the public. Improperly developed biodata, on the other hand, may present serious liabilities, both for the user as well as for those who may be directly affected. Consequently, would-be users of biodata are well-advised to seek and obtain any expertise they may lack, and to do so in advance of any operational use of biodata. Finally, public employers are encouraged to conduct and share biodata research to increase understanding and to help identify the characteristics of biodata that most contribute to its positive outcomes.

INTRODUCTION

Biodata (short for biographical data) is a name commonly used to refer both to a type of information as well as to the technology of how that information is used in employee selection. Usually, the information is obtained directly from job applicants via highly structured, multiple-choice questionnaires. Biodata typically include the kinds of information contained on job application blanks, in employment interviews, in personnel files, and in autobiographies. The technology of biodata is unlike most selection procedures in that the scoring of the information generally is not based on predetermined "correct" answers. Rather, the correct answers are based on how successful and unsuccessful employees answered the same biographical inquiries.

Recently, personnel measurement specialists and managers have begun to demonstrate an increasing interest in the use of biodata for selecting applicants into the public service (e.g., Sproule, 1990). This increased public sector interest in biodata can be attributed to several factors. Among these factors are: (1) the consistency with which biodata appear among the best predictors of future behavior, (2) biodata's absence or low levels of adverse impact by race or sex, while adverse impact continues to persist in many traditional tests of cognitive abilities, (3) the growing belief that the selection of job applicants can be improved by measuring more than just their cognitive abilities, and (4) recent advances in the development and the scoring of biodata. In addition, the *Uniform Guidelines for Employee Selection Procedures* (1978) and recent Supreme Court decisions (e.g., *Wards Cove Packing Co. v. Antonio*, 1989) continue to encourage employers to search for alternatives to traditional selection procedures.

Biodata[1] is not new. Biodata was first used in the life insurance industry to select successful sales personnel (Goldsmith, 1922). Mitchell (1987) estimates that approximately 75 percent of the life insurance sales force in this country and in Canada has been administered a biodata selection instrument. Besides the insurance industry, biodata has been used effectively in the private sector to predict a wide diversity of constructs. A partial list includes:

Job performance (Walther, 1961)
Employee turnover (Cascio, 1976)
Managerial effectiveness (Laurent, 1962)
Creativity (Buel, Albright, and Glennon, 1966)
Vocational interests (Mumford and Owens, 1982)
Student achievement (Freeberg, 1967)
Credit risk (Moran, Walsh, Clement, and Bumbeck, 1968)
Honesty (Rosenbaum, 1976)
Training success (Drakeley, Herriot, and Jones, 1988)
Career success (Childs and Klimoski, 1986).

The increased interest in biodata is reflected, in part, by the prominent inclusion of biodata in the U.S. Office of Personnel Management's (OPM) new Administrative Careers With America (ACWA) examination, an entry-level examination for federal administrative careers (Gandy, Outerbridge, Sharf, and Dye, 1989). Biodata also is being considered by the U.S. Army for possible use in the selection of military recruits (Wise, McHenry, and Campbell, 1990) and by the U.S. Department of Labor (Michigan Employment Security Commission, 1991) as a possible alternative to, or in conjunction with, its General Aptitude Test Battery (GATB).

---

[1]Generally, *biodata* will be used as a singular noun to refer to the biodata questionnaire or to the biodata technology. Sometimes, *biodata* will be used as a plural noun to refer to the biographical information and data obtained in the questionnaire.

Less well-known are the applications of biodata in various state and local jurisdictions. For example, biodata selection procedures are currently being used for clerical and welfare eligibility worker positions in California's San Bernardino County (French, 1991) and for supervisory positions in Kansas City (Dieckhoff, 1987). Biodata selection procedures have also been used for selections into police officer and related occupations (Richardson, Bellows, Henry, and Company, 1989) and are being made available by the International Personnel Management Association (1991) for the selection of police and correctional officers. In addition, personnel management specialists continue to successfully use biodata in the private sector. Unfortunately, due to their proprietary nature, most of these efforts tend to remain confidential and unpublished.

Despite the increased interest in biodata, the theory and technology of biodata are not yet well-formed, and those wishing to know more about this alternative selection procedure will find few organizations with biodata experience. Hammer and Kleiman (1988), in a survey of 718 personnel directors, found that only 15 (6.8 percent) of the 248 respondents had "ever used biodata," and only 1 (0.4 percent) was "currently using it." Hammer and Kleiman (1988) found that *lack of knowledge* (52 percent) was the second most important reason for not using biodata, second only to *lack of resources* (63 percent). Currently, much of the recent work and thinking on biodata can be found only in technical reports, dissertations, conference papers, and other media with limited public distribution. *The Biodata Handbook* (Stokes, Mumford, and Owens, in press) is expected to consolidate much of this information.

Furthermore, there are currently no readily-available cookbooks on how to develop a biodata selection procedure. England's (1971) book may be helpful, but it has been criticized for oversimplifications and for putting personnel selection in the hands of amateurs (Pace and Schoenfeldt, 1977). This criticism is not to be taken lightly, since personnel selection practices, as well as legal, technical, and practical requirements, have changed dramatically since the early 1970's.

PURPOSE

To help sort out some of biodata's complexities, this monograph outlines the potentially positive outcomes of biodata and the characteristics of biodata that seem to contribute most importantly to those outcomes. It also discusses some of the major issues and practical considerations associated with the development and use of biodata as an employee selection tool in the public sector. This monograph is designed to provide public personnel managers, policy makers, and selection specialists with a more complete understanding of the complex issues surrounding biodata, biodata's advantages, and its disadvantages.

This monograph does not purport to be an exhaustive review of the biodata literature or all the research conducted in the topic areas discussed. Rather, this monograph intends to highlight and summarize what we know about biodata as a personnel selection procedure and to focus on its use in the public sector. More detailed discussions on many of the topics included in this monograph may be found in the writings of Owens (1976), Schuh (1967), Freeberg (1967), van Rijn (1980), Barge and Hough (1988), Mumford and Owens (1987), and others.

The remainder of this monograph will: (1) define biodata and describe its possible role in public sector selection, (2) describe the potentially positive outcomes of biodata and the specific issues and challenges associated with each outcome, (3) discuss some general issues and considerations in the use of biodata, (4) examine the

characteristics of biodata questions that appear to be important contributors to biodata's positive outcomes, (5) outline practical concerns in the design and administration of biodata questionnaires, and (6) make general recommendations for future research and increased use of biodata in the public sector.

DEFINITION OF BIODATA

Despite the emerging interest in biodata, there has been relatively little systematic research to define biodata or to articulate the factors that are most likely to contribute to its validity and practical applications in public sector employee selection procedures. The factors contributing to biodata's validity are complex. Both the type of information and the technology of biodata are multi-dimensional and different measurement specialists have given different emphases to different components. Consequently, there is a great deal of confusion not only about what constitutes biodata, but also about which of its characteristics most contribute to its claimed advantages and disadvantages. These differences are evident in the variety of terms used to describe biodata: biographical inventory, life history data, biographical questionnaires, weighted application blanks, personal history, individual achievement record, life experience inventories, biographical information blank, and others.

With such a variety of definitions and approaches, it is virtually impossible to generalize about biodata. Owens (1976) notes that biodata is not one measure of one dimension but multiple measures of multiple dimensions. Results are often dramatically different; and, as Mael (1992) correctly points out, any claimed advantages or disadvantages for biodata may be true only for certain types of biodata.

Biodata, defined in its broadest sense, is a selection procedure based on information *about* the applicant. This information may include: (1) prior behaviors, activities, and experiences of the applicant (e.g., recreational, educational, or work experiences), (2) indirect indicators or *signs*

Figure 1. Topic Areas Historically Included in Biodata Questionnaires[1]

PERSONAL
- Age
- Marital status
- Number of years married
- Dependents, number of
- Birth order
- Physical health
- Time lost from job
- Size of home town
- Number of times moved
- Time at last address
- Nationality
- Weight and height
- Sex

BACKGROUND, GENERAL
- Occupation of parent
- Military discharge record
- Early family responsibility
- Parent family adjustment

EDUCATION
- Highest level of education
- Education level of spouse
- Major field of study
- Subjects liked, disliked
- Recency of education
- Grades, honors, awards

SOCIOECONOMIC
- Financial responsibility
- Number for creditors
- Loans as a portion of income
- Monthly mortgage payment
- Debts
- Net worth
- Amount of life insurance
- Properties/Investments
- Current living expenses
- Earnings expected

SKILLS
- Read/speak foreign language
- Read blue-prints
- Ability to type
- Repair work on cars
- Training for target job
- Possession of job skills
- Machines/tools/equipment

EMPLOYMENT
- Type of previous experience
- Worked while in high school
- Number of previous jobs
- Specific work experiences
- Self-employment
- Seniority
- Reason for leaving last job

SOCIAL
- Club memberships
- Attendance at group meetings
- Offices held
- Leadership experience

INTERESTS
- Preference for outside work
- Hobbies
- Sports
- Sources of entertainment

PERSONAL/ATTITUDINAL
- Willingness to relocate
- Willingness to travel
- Self-confidence
- Basic personality needs
- Drive or energy level
- Job preferences

[1]*Note.* Most of the topic areas included on this list would not be suitable or acceptable in public sector selection procedures, such as marital status, national origin, sex, family adjustment.

of achievements in these activities (e.g., awards, honors, grades), (3) information about the contexts or situations surrounding these activities (e.g., family income, parents' education level), and (4) personal information (e.g., age, marital status, health, skills, interests, and attitudes). Figure 1 shows some of the types of information traditionally included in biodata questionnaires. As will be discussed, *many of these traditional inquiries would be illegal or unsuitable for use in current public sector selection procedures.*

Biodata's focus on information *about* applicants distinguishes biodata from most testing procedures, such as assessment centers and written cognitive ability tests, that require applicants to *demonstrate* their ability. Biodata does not require applicants to perform.

Although the focus of biodata is typically historical, no technical requirement dictates which time period in a person's life is most suitable for biodata's inquiries. Moreover, in a direct comparison of "present-oriented" and "past-oriented" items, Kleiman and Faley (1990) found that "present-oriented" items performed as well or better than "past-oriented" items. Restricting biodata to either past- or present-oriented items might unnecessarily limit access to information that may be highly useful in the prediction of job success. Until more is known, there appears to be no compelling reason at this time to limit biodata to one type of item. Conceivably, a mixture of items may capitalize on the strengths of both item types.

Biodata's traditional focus on past behaviors is based, in large part, on the axiom in industrial psychology that "past behavior is the best predictor of future behavior." However, because other factors may also be relevant, such as

the type of behavior involved, its frequency, and the recency with which it was last performed, the focus on historically based information has not precluded questions, such as, "My peers would say that my ability to lead others is much above average...."

Although there are exceptions, which will be discussed later, biographical information typically has no a priori right or wrong answers. Rather, the scoring of biodata is usually derived directly from an external measure of job success. Biodata's scores are directly linked to measures of job success, not so much through any rational or judged linkages as through empirical, observable, and actuarial relationships. This is referred to as empirical scoring or empirical keying.

With the empirical scoring procedure, biographical information that is associated with a valid measure of job success is scored as "correct," and information associated with unsuccessful job performance is scored as "incorrect." A range of weights reflecting the degree of relationship with the job success measure is often used. Any information that fails to differentiate superior employees from barely acceptable employees typically is given a neutral weight or is deleted from the biodata questionnaire.

Empirical scoring is one of the ways in which biodata differs from traditional employment interviews, application blanks, and other selection procedures that also are based on information *about* applicants. While biodata is usually keyed empirically, these other selection procedures typically derive their scores from subjective judgments (at times pre-determined and unstandardized) about the relevance of the information for the requirements of the vacant position.

EMPLOYEE SELECTION IN THE PUBLIC SECTOR

To more fully understand the challenges faced by selection specialists in the public sector, it is important to know how employee selection in the public sector differs from that in the private sector. Some of the major differences between public and private sector employment practices were described in a monograph of the International Personnel Management Association Assessment Council (IPMAAC) prepared for the National Commission on Testing and Public Policy (Weisen, Abrams, and McAttee, 1990).

First, most public sector jurisdictions operate under merit principles, that require, in part: (a) recruitment, selection, and advancement of applicants or employees on the basis of their relative abilities, knowledges, and skills, (b) fair and equal treatment of all applicants and employees, without regard to their political affiliation, race, color, sex, national origin, religious creed, age, or handicapping condition and with proper regard for their privacy and constitutional rights as citizens. In the private sector, the merit principles of fairness and prohibitions against patronage are ideal goals. In the public sector, they are legally mandated.

Second, most government jurisdictions have laws that require open announcements of all vacancies. Such open announcements must state in detail the nature of the position(s) to be filled, the minimum qualification requirements, and details about the procedures that will be used to select applicants to fill the vacancies. The open announcements typically require public employers to develop separate and specific selection procedures for each type of vacancy.

For many vacancies, open announcements precipitate large numbers of applicants. It is not unusual to have hundreds or even thousands of applicants for some positions, such as police officer, firefighter, clerk, and welfare worker. While private sector employers need only find and select from a few good candidates, public sector employers must often evaluate large numbers of candidates and then select from among the many qualified candidates those who are most likely to succeed in the vacant position(s). The impracticality of most individualized assessment methods (e.g., oral interviews) for large numbers of applicants has led most public jurisdictions to rely heavily on machine-scorable examinations, at least as the initial method for reducing the number of applicants and identifying highly qualified candidates.

Third, public sector selection practices are often subject to close public scrutiny. This scrutiny derives not only from applicants vying for government jobs, but also from the general public, the news media, unions, and other interested parties. Each in their own way expects and demands that the government's selection process be straightforward, logical, fair, and open. Perceived or real unfairness, combined with the openness of the system, and the availability of careful records have made public sector selection practices likely targets for class-action lawsuits. This type of scrutiny is rare in the private sector, and its presence or absence substantially affects the way an employer's personnel selection system is designed and managed.

Finally, it is important to understand that government jurisdictions vary considerably in the laws, guidelines, and practices related to their personnel selection systems. Increasingly, state and local governments are prohibiting

many personnel practices not prohibited by federal laws (Mitchell, 1989). Sharf (in press) notes, however, that while states may prohibit certain practices, they may not authorize any employment practices that would violate federal law.

Clearly, public sector employers are faced with employee selection challenges far different from those of their private sector counterparts. The legal, technical, and practical issues are complex, and the stakes are high. Consequently, public sector employers are advised to retain appropriate professional expertise in the design and development of their employee selection systems.

BIODATA'S POTENTIAL OUTCOMES

This section describes the potential outcomes of biodata. For each outcome, we will first present some evidences for the potentially positive nature of the outcome. This will be followed by a brief discussion of the specific issues surrounding each outcome and some of the challenges that must be faced. Whenever appropriate, different approaches or possible options for achieving the positive outcome will be described. Generally *outcomes* refer to the effects of biodata in the selection system. Depending on the circumstances, identical biodata questionnaires may have different outcomes.

Not all potential outcomes are equally important. While some outcomes are technically essential (e.g., reliability and validity) or technically desirable (e.g., high utility and generalizability), others (e.g., low adverse impact and cost effectiveness) are more important for legal and practical reasons.

The order of presentation of the outcomes starts with the more essential technical requirements, moves through the legal requirements, and ends with a variety of practical considerations. Because many of the outcomes are highly interrelated, this order of presentation must not be rigidly interpreted to reflect either the relative importance of the outcomes or the emphasis to be placed on them during the design of a biodata questionnaire.

Throughout this section, it is important to view biodata as a complex multi-dimensional system that can be orchestrated in many ways to achieve a desired outcome. More than one combination of components may lead to a desirable outcome. Often, stressing the quality of one

component requires sacrificing another. Generally, the optimum combination is one that fits the particular needs, purposes, and contexts for which it was intended.

Validity

The one characteristic of biodata that stands out most conspicuously and is little challenged is biodata's high validity and ability to predict a wide range of criterion measures. In a previous overview, van Rijn (1980) reported that time after time, biodata questionnaires have proven to be one of the best predictors of future behavior. Asher (1972) found that 55 percent of 31 cross-validated validity coefficients were .50 or higher. In addition, McKillip and Clark (1974) and England (1971) showed that biodata proved to be a valid predictor of performance on jobs as varied as sales positions, scientists, office workers, middle managers, and military officers. More recent reviews confirm and extend these earlier findings (Barge and Hough, 1988; Mumford and Owens, 1987).

Even though the threefold distinction among validation models is no longer considered as useful as it once was (e.g., Landy, 1986), we nevertheless find it useful to discuss biodata in terms of these three models. However, rather than consider these models as distinct types of validity, they are best viewed as different types of evidence for different types of inferences that can be made from test scores.

Criterion-related validity. Meta-analyses comparing biodata against other predictor measures have been particularly useful in confirming biodata's potential validities. Consistently, these reviews found biodata to have criterion-related validities approaching and some-

times exceeding those for cognitive ability tests (Hunter and Hunter, 1984; Reilly and Chao, 1982; Reilly and Warech, 1990; Schmitt, Gooding, Noe, and Kirsch, 1984). A comprehensive review by Barge and Hough (1988) included over 100 published and unpublished studies conducted since 1960. Using a conservative strategy, they obtained median validities in the .20s and .30s for the prediction of training success, job performance, job involvement, and adjustment to work. Such validity coefficients, when corrected for criterion unreliability, may average .05 to .10 points higher (Reilly and Warech, 1990).

Content validity. Criterion-related validity has been the mainstay of biodata. Since empirical keying scores only items related to the measure of job success (i.e., the criterion), it is unlikely that the overt content and form of the scored biodata questionnaire will be an accurate sampling of the different aspects of job success. Moreover, a content domain would be difficult to specify for non-task measures of job success, such as tenure.

However, content validity of biodata may be a more appropriate and feasible strategy when biodata are developed specifically to reflect desired job behaviors or when biodata are rationally scored and weighted. One version of this approach was demonstrated by Mann (1983) to predict police officer turnover, while Pannone (1984) used biodata to predict performance on a written selection test, in lieu of traditional prescreens of education and experience.

Construct validity. Compared to content validity, construct validity is more compatible with the criterion-related validity strategy of biodata. Kavanaugh and York (1972), Quaintance (1981), and Russell and Domm (1990) found that theory-based biodata items tend to be particu-

larly effective in yielding sizable validity coefficients. However, Russell and Domm found that the theoretical constructs used to generate biodata items bore little relationship to those derived from a factor analysis of the biodata items after they were scored.

In a comparison study, Mitchell and Klimoski (1982) found that empirically-keyed biodata had significantly higher validity than rationally-keyed biodata, even after substantial shrinkage during the cross-validation. Although theoretical considerations help to yield useful biodata items, theoretical analyses may perhaps be most useful *after the fact* in helping to explain and further understand the biodata-criterion relationship once it has been clearly established.

With some exceptions, most efforts to establish biodata's construct validity (e.g., Owens and Schoenfeldt, 1979) have been in settings other than employee selection. Construct validity does not appear to be a prominent strategy in the validation of selection procedures. In an informal literature review, Freudenberg (1991) found that while construct validity often provided the grand context for test development, there was not one study that established an employment test as construct valid in the sense of the *Uniform Guidelines on Employee Selection Procedures* (1978).

For example, Gandy, Dye, and MacLane (in press), using a confirmatory factor analysis of OPM's Individual Achievement Record (IAR), identified four factors: work competency, high school achievement, college achievement, and leadership. Correlating these four factors against a battery of psychometric tests, resulted in a more complete descriptions and understanding of the constructs being assessed by the IAR. Based on the pattern of

intercorrelations, Gandy et al. concluded that the IAR measured both cognitive and noncognitive aspects of behavior. For a more detailed treatment of the construct validity research on the IAR, see Dye (1991).

Reliability

Reliability estimates of biodata, when they are reported, vary considerably. In a discussion of the reliability of biodata, Mumford and Owens (1987) noted that researchers (e.g., Plag and Goffman, 1967) have found biodata items to display relatively low intercorrelations with one another (i.e., low internal consistency reliability). This should not be surprising, since biodata questions typically cover a wide range of topics and only individual items related to the criterion are included in the final scoring. Nevertheless, biodata items that have conceptual and content similarity (e.g., high school and college academic achievement items) would be expected to have relatively higher intercorrelations. This was confirmed by Dye (1991), even though, high school and college achievements were found to be separate factors in the IAR.

Given the heterogeneous content of biodata, test-retest reliability estimates are more suitable measures of biodata's reliability than estimates based on internal consistency. Usually, the test-retest estimates of biodata reliability are quite satisfactory, ranging from the .60's into the .80's, prompting Mumford and Owens (1987, p. 10) to conclude that "background data items provide an unusually reliable description of differential behavior and experiences, even over relatively long intervals." Chaney and Owens (1964) found a test-retest coefficient of .85 on a 56-item biodata form that was re-administered after 19 months.

The combination of low item intercorrelations and high retest reliability also can be viewed as a distinct advantage of biodata questions. Mumford and Owens (1987) point out that the combination of these two characteristics permits just a few biodata questions to capture a great deal of descriptive information about the applicant.

Generalizability and Stability of Biodata Validity

One of the obstacles to the use of biodata is a concern about the generalizability of its validity. While research has amply demonstrated that the validity of *cognitive* tests generalizes across settings, organizations, and jobs (e.g., Lilienthal and Pearlman, 1983; Pearlman, Schmidt, and Hunter, 1980; Schmidt and Hunter, 1981), the evidences for biodata's generalizability are more limited and recent.

The advantages of generalizability are considerable. If biodata can be constructed to yield generalizable validities, organizations and jurisdictions that lack the resources and expertise to develop their own biodata scales would not be precluded from the use of biodata.

Many factors have been cited for the traditional lack of generalizability and stability of biodata. These factors include changes in: applicant groups, the nature of the target job, the definition of what constitutes success on the job, the labor market, workforce needs, and personnel policies. Poorly constructed items, small sample sizes and the particular vulnerability of biodata's scoring procedure to chance factors have also been cited. For example, Hunter and Hunter (1984) point out that since individual biodata items are selected solely because they correlate with the criterion, the procedure may pick up—purely by chance—items that only happen to work for that particular criterion or that particular sample of employees.

Since the risks of capitalizing on chance are particularly great when sample sizes are small, as was the case with many early biodata studies, it is not surprising that the generalizability and stability of biodata have historically been questioned (e.g., Anastasi, 1988). Generally, to help guard against chance factors operating in the development of biodata's scoring keys, cross-validation of empirical scoring keys is essential (Cascio, 1982). By using the scoring key developed on one group of individuals and applying it to another independent group, items that may only have been predictive by chance will likely manifest themselves. The greater the influence of chance factors, the greater the decrease in the validity of the biodata questionnaire for the cross-validation sample. Generally, the larger the validation sample, the less the drop in validity in cross-validation.

Although some early studies (e.g., Buel, Albright, and Glennon, 1966) suggested that biodata may be more generalizable than at first thought, it required recent evidence with carefully developed biodata questionnaires and large samples to expand on these earlier findings. First, Gandy et al.(1989) demonstrated that when a scoring key is developed across multiple occupations and organizations, the biodata validities are extremely stable. Their scoring key generalized to over 100 entry-level administrative occupations and 28 different federal agencies. The double cross-validation design with over 6,000 employees was considered an important factor in the stability of the scoring key.

While Gandy et al. demonstrated that a scoring key generalized to the occupations and organizations used in its development, Rothstein, Schmidt, Erwin, Owens, and Sparks (1990) showed that, contrary to conventional wisdom, biodata scoring keys *can* and *do* yield validities

that generalize across organizations that were not included in the original key development. In addition, they presented evidence that suggests that biodata generalizes across other variables frequently cited as moderators of biodata's validity: race, sex, education levels, years of company service, years of supervisory experience, and age. As in the Gandy et al. study, an important factor appears to be the strategy of basing the scoring key on large (10,000) heterogeneous samples from multiple organizations. Presumably, biodata questionnaires scored in this way capture the more generic and common elements of success that are important across all the jobs and organizations used in the scoring procedure.

In a discussion of the Rothstein et al. findings, Schmidt and Rothstein (in press) describe how meta-analyses of biodata to determine validity generalization present unique obstacles not typically encountered in meta-analyses of cognitive tests. Moreover, they note that these findings are particularly important, because they counter the arguments that biodata's sensitivity to situational and motivational factors may make biodata's validity especially susceptible to changes in such factors as organizational value systems, leadership styles, organizational culture, and management philosophies (Schmitt and Schneider, 1983).

Biodata scoring keys have also been criticized for eroding over time. For example, England (1971) reports that the predictive validity of a particular biodata questionnaire decreased over five years from a high of .74 to a low of .07. Contrary to this finding and popular belief, Brown (1978) was able to demonstrate that—with proper precautions—a biodata questionnaire could remain effective even after 38 years. He attributed his positive find-

ings to biodata questions that were maintained and kept current, to scoring keys that remained strictly confidential, and to scoring keys that were developed on very large samples of employees.

The studies of Gandy et al. (1989) and Rothstein et al. (1990) show that when biodata keys are developed across different situations, groups, or organizations, extremely stable keys can be developed. These studies also illustrate that there is no single validity estimate for biodata and that different validity estimates are likely to be obtained for different samples, different situations, and different measures of job success. Despite the promising finding that biodata validities need not be specific to particular organizations, additional research will be needed to identify further the factors that are most important in the generalizability and stability of biodata validities.

Adverse Impact

In addition to federal civil rights laws, various state and local fair employment laws prohibit employers from discrimination on the basis of race, color, sex, religion, or national origin. Although public employers are always required to use valid and fair employee selection procedures, they must be particularly prepared to demonstrate the validity or business necessity of their procedures when population subgroups are adversely impacted by the procedure.

England (1971), in the preface to his book on weighted application blanks, wrote that the nature of the items used and the techniques of item weighting make it quite probable that a *weighted application blank* (i.e., biodata) developed on one cultural group will discriminate against any individual from a group with quite different biographi-

cal background. However, this ominous warning tends not to reflect current findings. It more accurately applies to the strategies of early developers of biodata who rarely prescreened biodata items. Biodata questions that today would be screened out because they were considered discriminatory, invasive, or otherwise offensive to one or more subgroup were not uncommon in earlier and more traditional biodata questionnaires. Unfortunately, as reported in England (1971), many of these items were highly predictive of job success across a variety of occupational groups.

Even in the absence of item prescreening in earlier biodata questionnaires, Reilly and Warech (1990) still found that ethnic group differences reported in the earlier studies tended to be smaller than those found on many cognitive ability tests. Generally, major ethnic group differences have not been found in studies for which subgroup data were reported (e.g., Cascio, 1976; French, 1986; Gandy et al., 1989; Hough, 1984; Lefkowitz, 1972; and Ritchie and Boehm, 1977).

Relatively fewer studies have investigated sex and age differences. While Nevo (1976) and Webster, Booth, Graham, and Alf (1978) found male-female differences, others (e.g., Gandy et al., 1989; Hough, 1984; and Owens, 1971) found few major sex differences. Similarly, while Laurent (1962) and Ferguson (1967) found some age differences, Gandy et al. (1989) did not.

As with all the potential outcomes of biodata, the extent to which a biodata questionnaire impacts adversely on a applicant subgroups seems to depend, in large part, on the particular types of questions used. By carefully avoiding items with potential adverse impact and/or prescreening

items with such an effect, adverse impact can be dramatically reduced. Mumford and Owens (1987) cite several studies in which rational prescreening of item content minimized objectionability and bias against certain subgroups, such as women or minority group members.

Although item prescreening risks lowering the criterion-related validity of the biodata questionnaire, the potential practical benefits to be derived from the trade-off can be considerable. By removing 20 of 64 items, Gandy et al. (1989) were able to reduce the black-white subgroup effect size (i.e., subgroup difference) from .34 standard deviations to .26 standard deviations, the Hispanic-white effect size from .20 to .08, and the male-female effect size from .26 to .15. The cost for these substantial reductions in subgroup differences was a minimal decrease in the predictive validity (i.e., correlation coefficient) for the total group — from .33 to .32.

Alternative strategies for reducing adverse impact include counterbalancing the items with adverse impact or developing separate scoring keys for different subgroups. The first strategy involves balancing biodata items in such a way that all subgroups will be equally affected by items with adverse impact. Theoretically, the adverse effects cancel themselves out over the different subgroups. The second strategy requires large samples of employees for each subgroup and is designed to capture criteria for all of the subgroups in the best possible way.

Neither of the preceding two strategies is recommended for reducing adverse impact. The first may not remove all offensive items from the questionnaire and is likely to precipitate negative applicant reactions. The second strategy, although technically feasible (e.g., Thayer, 1977),

presents practical, social, and ethical concerns and would likely violate the prohibitions against within-group norming of the 1991 Civil Rights Act.

Relatively little research is available to guide the practitioner in the prescreening of items to reduce adverse impact. It may well be impossible to accommodate all subgroups. However, items specifically asking about issues of race or sex are clearly prohibited. Similarly, it is best to avoid items that may be indirect indicators of subgroup membership, such as place of residence, ownership versus rental of home, family income, parents' education level, and others. Other questions are best avoided for ethical, social, and privacy reasons. These include questions that perpetuate subgroup stereotypes, that may be offensive to subgroup members, that are considered unnecessarily invasive of the applicant's right to privacy, or that pertain to circumstances or situations in the past over which the applicant had little or no control. Pretesting potential biodata questions with individuals from relevant applicant subgroups may be a useful way of identifying offensive items. Such sensitivity screening is increasing as an important component in many test development programs.

Fairness

Analyses for fairness typically require a comparison of applicant subgroups to ensure that the predictor does not unfairly over- or under-predict job performance. Psychometric debates about fair selection models have abated, and the "consensus seems to be that (a) the Cleary model is the appropriate test for bias and (b) bias is shown by differences in the regression slopes" (Guion and Gibson, 1988, p. 356).

Using regression and chi square procedures described by Gulliksen and Wilks (1950), Gandy et al. (1989) demonstrated that a biodata questionnaire developed to meet both technical and public sector requirements can be a fair predictor of job success for all major subgroups. Basically, they found no subgroup differences in standard error or regression slopes, but the intercept differences indicated a small amount of overprediction for blacks.

Nevertheless, selection specialists need to balance a variety of conflicting forces. Although psychometricians may have reached consensus on the issue of fairness, the legal and social issues are far from resolved. For example, the courts still seem to think of group differences alone as evidence of bias (Guion and Gibson, 1988), and Kleiman and Faley (1985) note that judges often sidestep technical issues and apply their own views of commonsense.

Incremental Validity

The relatively low correlations of biodata with most other predictor measures (Reilly and Warech, 1990) and biodata's high correlations with measures of job success make biodata a particularly attractive component for a composite predictor measure. Booth, McNally, and Berry (1978) found that the addition of just three biodata items to a cognitive aptitude battery increased the multiple correlations with training performance from .35 to .48. Generally, the validity of a composite predictor may be expected to increase to the extent that the composite's components measure different and important aspects of job success. For example, the addition of non-cognitive, temperament items to the U.S. Army's present cognitive ability battery substantially improved prediction of job performance. As expected, the largest increases were in

those areas of job success, namely *leadership* and *discipline*, where the cognitive predictor was relatively weak (McHenry, Hough, Toquam, Hanson, and Ashworth, 1990).

Despite biodata's potential as a component in a composite selection procedure, preliminary validity evidence for OPM's ACWA examination demonstrated only a modest increase in total validity for the composite selection procedure (U.S. Office of Personnel Management, 1990). However, it remains to be seen in the follow-up applicant validity study to what extent this effect persists.

The incremental strength of biodata may derive from biodata's sensitivity to situational, extrinsic, and other non-personal factors associated with past achievements. While cognitive ability tests focus on the capabilities of a person and are relatively insensitive to situational variations, biodata seems to reflect more of the person-situation interaction, that is, what the person has achieved, given his or her capabilities *and* particular situation or environment. Hunter and Hunter (1984) suggest that biodata may be capturing more of the *will do* aspects of job performance, while more traditional cognitive measures tend to predict what applicants *can do*. By capturing more of what a person has *typically* done in the past, greater prediction will be possible for what a person will *typically* do in the future (Mitchell, 1989)

Legal Defensibility

Sharf (in press), in a search of cases brought under Title VII of the Civil Rights Act, found no challenges to the use of biodata in employee selection. Sharf did note, however, that biodata was upheld under one state law and that the

Department of Justice recommended a biodata instrument in one public sector consent decree. Although this paucity of legal precedent for biodata may be attributed largely to the relatively infrequent use of this procedure in the public sector, it may also be attributed to biodata's high levels of validity and low adverse impact. Selection procedures with these two characteristics do not usually precipitate legal challenges.

However, high validity and low adverse impact alone do not ensure biodata's legal defensibility. Hunter and Schmidt (1976) point out that some biodata items might be challenged for invasion of privacy or might be objected to as indirect indicators of race or sex. As van Rijn (1980) noted, current legal guidelines, statutes and regulations have placed considerable constraints on the types of questions permitted in pre-employment inquiries, particularly in the public sector. These constraints generally are imposed to protect applicants from unwarranted invasions of privacy and employment practices that deliberately or inadvertently deny employment on the basis of sex, race, national origin, religious preference, or other non-job-related factors, such as marital status, home ownership, and age.

Despite their potential validity, employers are advised to consider each biodata inquiry carefully. Although there are no clear standards or guidelines, employers should make every effort to ensure that there are no unwarranted invasions of privacy, that an applicant's reasonable sensitivities are not offended, and that the inquiries are not likely to cause adverse impact by race, sex, national origin, age, religion, or handicapping condition. Questions probing sensitive areas of personality, motivation, sexuality, and interests, in particular, are generally best avoided.

Other questions, such as education level of the applicant, may be unwarranted for some jobs, such as unskilled laborer, yet be acceptable for other jobs, such as engineering.

Mitchell (1989) and Sharf (in press) note that many states have independently passed a wide variety of regulations and guidelines restricting the use of many types of biographical questions. These restrictions are in addition to the relevant federal requirements of Title VII of the Civil Rights Act of 1964, as amended in 1972, the Privacy Act of 1974, and now the Civil Rights Act of 1991.

Arvey (1988) and Ash (1988) have identified many of the pre-employment inquiries that are most vulnerable to legal attack in various states. Mitchell (1989) observes, however, that these state guidelines generally were intended to apply to application forms or employment interviews. Consequently, it remains unclear to what extent any of the prohibited questions might be permissible within the context of a professionally developed and job-related biodata selection procedure, since most states explicitly allow the use of professionally developed selection tests. Complicating the issue further, Mitchell (1989) points out that most of the state prohibitions are not themselves laws or statutes. Rather, they are often characterized as guidelines, interpretations of the law, or merely opinions of a state commission—having the force of neither law nor precedent.

Pace and Schoenfeldt (1977) assert that while the criterion-related foundation of biodata, at face value, may be entirely permissible from a legal standpoint, it may violate the spirit of the Civil Rights law. They argue strongly against the use of any items that bear no *rational*

relationship to the job in question and suggest that biodata include only items previously identified as reasonable on the basis of content considerations.

Crosby and Mitchell (1988) suggest that Pace and Schoenfeldt's directives may be overly restrictive. They note that the *Uniform Guidelines On Employee Selection Procedures* (1978), published a year after Pace and Schoenfeldt's article, made it clear that criterion-related validation, in and of itself, was a principle means of establishing the job-relevance of a selection procedure. Nevertheless, the American Psychological Association (1988), in an *amicus curiae brief* for *Watson v. Fort Worth Bank and Trust*, reiterated the views of Pace and Schoenfeldt.

Despite Crosby and Mitchell's (1988) argument, current selection specialists are well-advised to supplement biodata's empirical relationships to job success with rational linkages. Does this mean that all biodata items must always be theoretically interpretable? There are no clear answers, and even advocates of theoretically interpretable selection procedures (e.g., Guion, 1965) admit that it would be a poor decision for a selection specialist to refuse to use something he or she did not fully understand, when it adds appreciably to the overall validity of a selection procedure.

In the absence of clear-cut guidelines, many decisions about which items are acceptable or unacceptable will need to be made on a case-by-case basis. Selection specialists need to be aware that for all their adherence to technical and legal requirements, some "judges prefer to prejudge the validity of the examination solely on the basis of an assessment of the content of the test....and to sidestep technical issues and rely instead on a commonsense approach to validation" (Kleiman and Faley, 1985, pp. 809, 813).

Acceptability

A practical desired outcome of any employee selection procedure is that it be considered acceptable to reasonable applicants. Davey (1984) considers this area one of the "fuzziest" in employee selection. However, because the acceptability of the selection procedure directly influences the likelihood of legal challenges, this may often be the overriding factor in the decision "to test or not to test."

Acceptability to applicants involves many of the issues discussed under legal defensibility. In addition, it is important that applicants generally perceive the selection procedure to be fair, job-related, and non-intrusive.

To be perceived as job-related does not necessarily require that each biodata question closely reflect the behaviors of the target job. It *does* mean, however, that applicants need to be convinced that the questions asked are related to the target job. Educational efforts to explain how biodata questions are empirically linked to job success may be one way to achieve this objective. Introducing an operational biodata questionnaire in terms of a structured application blank or a "paper-and-pencil" interview may help overcome some of the initial negative reactions applicants may have towards this "new" type of selection procedure.

To make increased use of biodata's full potential, educational efforts that explain to applicants how biodata are empirically linked to job success may eventually increase the number of biodata questions that will be considered job related and acceptable for inclusion in the selection procedure. Non-offensive neutral questions that on the surface

do not appear job-related may come to be acceptable once applicants more fully understand how responses to these questions are predictive of job success.

There is some limited support for the acceptability of the types of inquiries found on biodata questionnaires. Johnson, Newton, and Peck (1979) observed that most applicants have an expectancy that they will be asked to fill out an application blank and they usually perceive the request for biographical information to be a legitimate action. In another study, Stone, Stone, and Hyatt (1989) found that in a direct comparison of 12 (hypothetical) employee selection procedures, application blanks were rated *least* invasive by a sample of 84 adult volunteers. In addition, OPM's new biodata questionnaire seems to be well-accepted, and there have been no negative public reactions to this "new" selection procedure.

More studies of the factors contributing to the perceived invasiveness or acceptability of pre-employment inquiries would be beneficial to the design of future biodata questionnaires. Mumford and Owens (1987) cited several studies which indicated that studies of item acceptance can be useful. However, because society's definitions of objectionability and stereotyping are constantly changing, it is important to have current information about these factors in prescreening items for inclusion in a biodata questionnaire.

It is of interest to note here that Hammer and Kleiman's (1988) survey of personnel officers found that negative attitudes towards biodata were not generally considered important factors for biodata's non-usage. Only *equal employment opportunity risks* (46 percent) and

invasion of privacy (40 percent) were rated—by a significant portion of the respondents—as important impediments to biodata's usage. *Low face validity, unsuccessful past experience with biodata, questionable validity, high fakability,* and *unstable scoring weights* were considered important reasons for non-usage by fewer than 30 percent of the personnel officers.

The length of time required to administer the selection procedure is also an important consideration in the acceptability of the procedure. Although it may not seem inappropriate to require applicants to devote three or more hours to an examination that may determine their life's career, pressures from applicants and the general public are pushing towards shorter and less threatening examinations.

In an effort to be responsive to the needs of the general public, OPM has shortened its ACWA examination from 3-1/4 hours to 1-1/4 hours. The number of questions on the biodata portion of that examination was reduced from 112 to 46, with minimal effects on its validity and adverse impact (MacLane, 1991). Like most biodata questionnaires, both the long and the short form of the biodata questionnaire contained items that are not scored. Unscored items are experimental items or are used to detect faking and other response distortions.

Practical Feasibility and Cost Effectiveness

Practical feasibility and cost-effectiveness are relative terms. While biodata may have been practically feasible and cost-effective for use with the hundreds of thousands of applicants for federal administrative jobs, the circumstances under which these conditions occur may be relatively rare for most jurisdictions.

First, as will be discussed more fully later, empirically-keyed biodata requires access to large numbers of employees. Requirements for 400 to 1,000 employees are not uncommon. Not surprisingly, the extent to which sample size is a problem is a function of the size of the organization or jurisdiction.

Hammer and Kleiman (1988) found that sample size was a problem for 43 percent of the employers with less than 500 employees, for 28 percent of the employers with 500 to 999 employees, and for only 5 percent of the employers with more than 10,000 employees.

Similarly, biodata's structure, multiple-choice format, and objective scoring procedures make biodata particularly well-suited for assessment of the large numbers of applicants often encountered by public employers. Biodata are easily collected at the time other selections procedures are administered. The additional costs involved would be minimal. Biodata is most cost-effective when administered to and used to screen large groups of job applicants.

A third factor in the feasibility of biodata is the access to specialized expertise. This usually requires the expertise of an industrial/organizational psychologist or other selection specialist who is familiar not only with biodata technology but also with the many relevant current legal, social, and practical constraints that apply to public sector selection practices.

Although biodata is more feasible and cost-effective than various individualized assessment procedures, biodata, like other valid selection procedures for large applicant groups, requires considerable financial resources. One of the more significant costs is the time of the hun-

dreds of employees that are required to develop the biodata questionnaire. Depending on the criterion to be predicted, costs can vary from the relatively low, as in the prediction of employee turnover, to costly, as in the case where a complex measure of job performance must be developed. Not the least of the costs are those for the technical expertise that is required to develop a sound biodata questionnaire and the time and resources needed to do the statistical analyses to select questions, remove adverse impact, key items, and cross-validate the scoring key.

Gandy et al. (1989) roughly estimated that it cost $468,000 to design, develop, and key OPM's biodata predictor. However, this cost also contributed to the validation of OPM's cognitive predictor battery and is not unlike costs typically incurred for such large-scale validation efforts (Outerbridge, 1979). The magnitude of the cost, to a large extent, was directly related to the large number of research participants that were used in this study, namely, over 13,500 employees and 13,500 supervisors.

Despite the costs, there are also benefits to be derived from the use of valid selection procedures. Hammer and Kleiman (1988, p. 90) wrote that "[c]oncern about the cost of implementing biodata may be misguided; the benefits [of biodata] may far outweigh the costs." In a not so recent analysis of costs, Lee and Booth (1974) calculated that using biodata to hire 200 long-term clerical employees amounted to a cost-saving of $250,000 over a 25-month period. This suggests that even for relatively low-level jobs with relatively modest hiring and training costs, dollar savings from the introduction of a valid biodata procedure can be quite large.

The benefits of valid selection procedures can be considerable. This is largely because the benefits are ongoing. The improvements in increased productivity do not just occur once, they persist and continue each year the new employee continues to work for the organization. If the selection procedure improves tenure, considerable cost savings occur in the areas of recruiting, selection, and training.

The ongoing benefits of higher productivity and increased tenure are not the only benefits to be derived from a valid selection procedure. Selection procedures that are technically sound, legally defensible, and that have little or no adverse impact are less likely to incur the costs of legal challenges. However, these types of cost savings are particularly difficult to document and estimate.

Costs for biodata can be reduced considerably if a jurisdiction uses existing and previously validated biodata questionnaires. Rothstein et al. (1990) demonstrated that, like cognitive ability measures, carefully developed biodata are generalizable across organizations. However, the generalizability of such previously validated biodata questionnaires will most likely depend on the similarities between the occupations and organizations involved.

Accuracy of the Information

Since biodata usually involve self-reports of information about an applicant's past, there is potential that the information provided may be inaccurate. The accuracy of biodata information is an important concern (e.g., Anastasi, 1988; Fleishman, 1988) and a major consideration in the design of a biographical questionnaire.

susceptible to the effects of social desirability responding than biodata items that were scored so that each response option had its own unique weight (option-keyed) based on its relationship to the criterion.

The greater susceptibility to faking of item-keyed biodata was attributed to the fact that respondents need only guess the right direction (positive or negative) of the item's relationship to the criterion and then respond with an extreme response to obtain the maximal score. Option-keyed biodata only credits some of the response options, and extreme scores may or may not receive credit. Although linear relationships are possible in option-keyed items, this tends not to be the case. Typically, the five alternatives might be scored, as follows: $a = 0$, $b = -1$, $c = 0$, $d = +1$, and $e = 0$. Clearly, without knowing the pattern of response scores, it would be difficult for an applicant to obtain a higher score by faking such option-keyed items.

Little is currently known about how much distortion occurs or how it affects the outcomes of biodata. Many of the previous findings were derived from laboratory experiments rather than operational selection procedures. However, until more is known about the magnitude of faking in real operational settings, concerns about faking, in and of themselves, should not preclude the use of biodata.

To help reduce possible faking, employers are encouraged to use item types resistant to faking, to include instructions that the biographical information provided may be verified, and to apply special "lie-detector" scales or other techniques for identifying potentially inaccurate responses (e.g., Hough, et al. 1990). Mitchell (1987) lists a number of techniques to help reduce faking and other

falsifications of biodata. First, certain identical or related questions might be repeated and the answers compared for consistency. Second, lie-response or rare-response questions might be embedded within the questionnaire. These types of questions ask respondents about experiences that are either impossible or highly improbable.

Mitchell also discusses some techniques more suitable for biodata questionnaires presented via a computer. Among these techniques are: (1) random administration of biodata questions to confound applicants applying a predetermined key or template; (2) adaptively branching and tailoring biodata questions to each applicant; (3) monitoring response times for inconsistencies; (4) comparing response profiles for identical response patterns; (5) encrypting responses to limit access to high-scoring profiles; and (6) asking personal identification questions, e.g., mother's maiden name to identify stand-ins.

Clearly, biodata can be falsified in a variety of ways. However, the use of verifiable questions, lie-detection techniques, and instructions that the answers provided might be verified may help to reduce faking. In addition, the instructions might emphasize the need for accurate responding, so that applicants will obtain more precise feedback about their likelihood of success on the target job.

Finally, to further help reduce faking, biodata questionnaires as well as their scoring keys are best kept strictly confidential. In this way, biodata is more like a traditional written test than an application blank. The objective of not making the questions generally available is to reduce advance "gaming" of the answers by would-be applicants or by commercial vendors of services to improve selection test scores.

Flexibility

Finally, biodata offers a very flexible alternative to written cognitive ability tests. Biodata can be used in a number of ways: (1) to assess different aspects of the job applicant, (2) to measure the same aspect but in a different way, or (3) to assess some combination of the two. Simultaneously, biodata can be designed to predict the same or a different measure of job success from that predicted by a cognitive test.

Because biodata is well-suited to measuring different aspects of the job applicant and can be used to predict a wide variety of job success measures, biodata is a particularly attractive alternative or supplement to traditional selection procedures. There is currently a growing interest in considering not only the "whole applicant" (cognitive and non-cognitive) in making employee selection decisions, but also in defining job success more broadly than performances on job tasks (e.g., Pulakos, Borman, and Hough, 1988). Although cognitive ability tests are effective predictors of training and job task performances, there is increasing acknowledgement that job success is multidimensional, that it is not one thing, and that success often constitutes more than just the amount and quality of work performed, such as *how* the work was performed. Weisen et al. (1990) identified the measurement of the affective domain (personality and interests) as one of two more promising psychometric research areas.

Helmreich, Savin, and Carsrud (1986) illustrated the multidimensional nature of job success in routine jobs. They found that, initially, while employees were learning a routine job, the employees were limited primarily by their own abilities and experience. However, once they had

learned the job and passed the so-called "honeymoon" period, their subsequent performance became increasingly a function of their motivation. Biodata may be a particularly useful way of capturing these more long-range and non-cognitive aspects of job success.

Although biodata are not usually used to predict performance on other predictor measures, there is nothing to prevent biodata from being developed to predict performance on a cognitive ability test. There is some research to demonstrate that empirically developed biodata questionnaires can be effective predictors of cognitive tests (e.g., Olson, 1981; Pannone, 1984; Sparks, 1965) as well as of assessment center performance (Quaintance, 1981; Ritchie and Boehm, 1977). In such cases, biodata can become an alternative way of measuring cognitive performance or possibly a prescreen for a predictor that may be particularly difficult or expensive to administer.

As another example of biodata's flexibility, Myers and Errett (1959) first used a set of 19 biodata questions to predict who should be hired. Subsequently, they used the same set of questions to differentiate employees who were terminated from those who were not. Two quite different scoring keys were developed. This highlights how biodata can answer different selection questions and make different predictions based upon how it was developed and scored. Kluger et al. (1991) found that a biodata questionnaire scored by both item-keying and option-keying approaches may be capturing different aspects of the criterion domain. They suggest that different biodata keys for the same biodata questionnaire might be combined to increase the overall predictive validity of the questionnaire.

MAJOR ISSUES AND CONSIDERATIONS

Despite the appeal of biodata as an alternative to cognitive ability tests, there are some significant concerns and issues that need to be considered and addressed before biodata can be adopted as a selection strategy in the public sector. These concerns and issues include methodological, legal, ethical, and practical challenges. Unfortunately, there is relatively little research in the public sector to guide employers who may be interested in biodata as a potential selection procedure.

Scoring Procedures

One of the major debates regarding biodata centers around the way answers on the individual questions are summarized into a final score. Mumford and Owens (1987) outline four basic techniques: (1) empirical keying, (2) rational scaling, (3) factorial scaling, and (4) subgrouping. Each has advantages and disadvantages.

Empirical keying. The empirical keying approach is the dominant and traditional approach to the scoring of biodata. It has the important advantage of being directly related to an external measure of job success, and it is typically designed to maximize the prediction.

The *vertical percent* method advocated by England (1971) is one of the most common approaches to empirical keying. It assigns weights to responses based on the response differences between high and low performers. There are many other weighting approaches, each with advantages and disadvantages. Comparisons of different

empirical scoring procedures can be found in studies by Aamodt and Pierce (1987), Lecznar and Dailey (1950), Telenson, Alexander, and Barrett (1983), Malone (1978), and others.

Whatever empirical scoring approach is used, its adequacy depends largely on the adequacy of the job success measure and on the reference group used to establish the key. For example, Brumback (1969) cautions that if biodata is being used to predict a supervisory rating of job performance, the latter may be contaminated by a variety of factors, including discriminatory stereotypes.

Crosby and Mitchell (1988) view the use of an external criterion as a real improvement over earlier logical-content approaches to psychological measurement. These earlier approaches, now largely discredited, involved asking subjects how they viewed or evaluated themselves on a variety of traits or characteristics. Their answers were presumed to be true. This had two drawbacks. First, subjects were not always good self-evaluators, and second, they could readily distort their responses to enhance their standing, particularly when a job was at stake and the content areas of the inquiries were obviously related to the target job.

Although the direct link of biodata to a criterion measure is one of its greatest strengths, it is also one of its most pervasive difficulties. The difficulties do not derive so much from the criterion itself as they do with the way the link is established. Because items are selected solely on the basis of their ability to predict complex measures of job success, the item content of biodata is often heterogeneous and lacking in psychological meaningfulness. That is, the

complexities of biodata are not readily reduced to simple explanations, and theoretical explanations about how or why individual items predict job success are usually elusive.

Researchers, not without reasons, tend to abhor "blind empiricism" and typically prefer to have full and complete understanding of all the relationships among variables. However, the complexities of human behavior seldom afford personnel practitioners such luxury. Even models to predict a behavior as seemingly straightforward as employee turnover involve dozens of variables in all sorts of complex interactions (e.g., Mobley, 1982).

The strongest argument for empirical keying is that *it works*. It predicts job success and, as such, is job-related. Not knowing or not fully understanding all the complex underlying psychological relationships, in and of itself, may not be a sufficient reason for rejecting an empirically-keyed biodata questionnaire in the public sector. However, as will be discussed later, factors such as the large numbers of individuals required for empirical keying may limit the use of this approach.

Rational scaling. Rational approaches to scoring biodata are a major alternative to empirical keying. Mumford and Owens (1987) discuss two strategies: direct and indirect. In the direct approach, the behaviors that differentiate superior from barely acceptable performance are described and biodata items are developed to obtain evidences for these behaviors. In the indirect approach, the differentiating job behaviors are first related to psycho-

logical constructs thought to underlie the job behaviors, e.g., ability to interact with people. Then, biodata items are developed to assess the intermediate underlying psychological constructs.

While the indirect approach tends to result in items that have little apparent correspondence to the measure of job success, the approach is designed to yield greater construct validity and greater psychological meaningfulness. Like the empirical approach, however, these items may be objectionable to applicants, because the linkages between the items and the job success measure are not apparent. The direct measures, with their greater face validity, tend to have greater content validity and be more acceptable. However, they are also likely to be more vulnerable to falsification.

In one variation of the rational approach, Matteson (1978) used a common set of biodata items to develop 12 different rational scoring keys based on subjective judgments concerning the item content. The resulting keys were both homogeneous and relatively independent, and they compared favorably to an empirical key in predicting performance on refinery operating jobs. However, as expected, the specific nature of the single rational keys did not achieve the level of validity of the more heterogeneous empirical key.

Mann (1983) took a different approach. Working with job experts and using an analysis of the police officer job, he identified eight experiences hypothesized to be precursors of police officer turnover. Simple "yes-no" biodata items were developed to obtain information about each experience. The biodata items were designed to be content

JOEL P. WIESEN, Ph.D., DIRECTOR
APPLIED PERSONNEL RESEARCH
6 CAULFIELD CIRCLE
NEWTON CENTRE, MA 02159

valid, to have high face validity, and to be verifiable. An example of one item is, "Did you ever have a job for which you had to be outside in all kinds of weather?" A concurrent validity study suggested that the eight scored biodata items predicted police officer turnover over a period of two and one-half years.

While Matteson and Mann used rather direct content and face valid approaches, Mitchell (1991) used a more indirect approach to predict carefully-developed ratings of *learning rate, teamwork,* and *turnover* in three different high technology manufacturing jobs. Instead of developing and scoring biodata items directly from a job analysis, Mitchell, with the assistance of subject-matter experts, first identified the biographical antecedents that underlie important aspects of job success. These antecedents then became the logical bases for biodata questions and an a priori scoring procedure. Because the items were not directly linked to the job, they were not as transparent and did not have the same limitations as more face valid items.

The promise of Mitchell's a priori rational approach is evident in the obtained, uncorrected, average validity coefficient of .29. Mitchell notes that a major advantage of a priori rational scaling is that biodata questions can be developed and then validated in much the same way as traditional paper-and-pencil cognitive tests. Because the validity is built in, Mitchell's biodata questionnaires should readily generalize to occupations with profiles of antecedent behaviors similar to those of the occupation used in the development of the key. This, plus the reduced need for large samples sizes in the development of an empirical scoring key, opens up the application of biodata to jurisdictions that would not otherwise find biodata feasible.

Factorial scaling. Mumford and Owens (1987) point out that both empirical and rational scaling procedures have close ties to specific criterion measures. Consequently, different scoring keys may be necessary for different criteria and different situations. The factor analytic approach seeks to identify the summary dimensions underlying the universe of biodata inquiries.

Factor analyses are based on the correlations among individual biodata items. The correlations in the matrix can be derived in a number of different ways. First, the individual biodata items may be scored rationally or empirically. Second, individual biodata items may be item-keyed (i.e., each item consists of multiple responses on a scored continuum) or option-keyed (i.e., each response option enters the correlation matrix as a scored binary item). Third, the correlation matrix may be derived from a combination of these approaches.

In a direct comparison of the predictive power of empirical keys and factorial scales, Mitchell and Klimoski (1982) found that empirical keys were more predictive of job success than factorial scales, in both the original derivation sample and the cross-validation sample. The factorial scales, however, yielded more psychologically meaningful dimensions and were more stable during the cross-validation.

Although there is some promise to this approach (e.g., Baehr and Williams, 1967; Childs and Klimoski, 1986; Gandy et al., in press; Owens and Schoenfeldt, 1979), it is not without its disadvantages. First, like the empirical approach, it requires large sample sizes. Secondly, the factors to be derived depend to a large extent on the items

included in the analysis. Analyses of items keyed empirically against a criterion would yield different underlying structures than analyses of items keyed rationally. The issue is further complicated because the scoring of the items used in the factor analysis can vary. Each different way of keying the items, whether empirical or rational, could theoretically result in a different factor structure. For example, biodata items keyed empirically to a turnover criterion would not necessarily yield the same factor structure as items keyed to a measure of training.

While factor analyses do not appear to be particularly independent of situation-specific considerations, they are a useful tool in identifying the underlying psychological structures of both empirically keyed and rationally scaled biodata questionnaires.

Subgrouping. The fourth and final approach to the scoring of biodata is subgrouping (Mumford and Owens, 1987). This technique is wholly independent of any criterion measure or of specified populations. Rather, this approach seeks to identify modal patterns of background experiences and developmentally significant behaviors. All individuals are classified according to their developmental profiles. Although preliminary evidence suggests that this approach has promise as a general predictive system (e.g., in vocational counseling and social development), there remain many unresolved questions about the use of such an approach in the merit system context of the public service. Among these questions are questions of fairness and job-relatedness.

Summary of scoring procedures. Clearly there is a great need for more research on the different scoring procedures, particularly in public sector employment settings. While each of the approaches has advantages and disadvantages, there is much that we do not know. More importantly, except for subgrouping, the scoring approaches are not mutually exclusive and each may include components of the other. For example, rational scaling may overlay an empirical key, and factor analysis of a scored biodata questionnaire can add substantially to our understanding of what is being measured.

In a discussion of the issue, Barge and Hough (1988, p. 97) wrote: "[w]hile the increased conceptual orientation is laudable, it is noted that such effort accrues much of its value through previous empirically demonstrated validity." They argue, moreover, that the absence of an empirical base should not deter public sector employers from beginning the process of building an empirical data base from which theories can then be developed.

Finally, until the courts more fully clarify empirical "job-relatedness," employers are advised to strive to develop biodata questionnaires that not only predict the criterion, but which also *look* like they would predict. If the selection procedures do not *look* job-related, logical and understandable, the potential for challenges from applicants and others will be high. Although specific scoring keys are best not made known to applicants, some well-documented examples and explanations of the underlying relationships between the selection procedure and the criterion may be useful.

Concurrent Versus Predictive Validity

An important validity issue concerns the use of concurrent versus predictive designs. Briefly, in concurrent designs, the predictor and job performance measure are simultaneously obtained on *job incumbents* and then compared. In predictive designs, the predictor is typically administered to *job applicants* for a position and is subsequently compared to the job performance of those applicants who were hired. Ideally, from a validation point of view, all job applicants would be hired in this approach. However, few organizations could afford to hire all applicants, regardless of their level of ability or skill.

Each validity strategy has its advantages and disadvantages (e.g., Arvey and Faley, 1988). While the predictive design provides a more accurate measure of *applicants*, it may also be costly and time-consuming. The concurrent strategy, on the other hand, risks the possibility that incumbents will differ significantly from applicants and that their predictor performance will yield imprecise estimates of how applicants might perform on the predictor. Simply having worked in the occupation for a certain length of time, for example, may affect how incumbents respond to some biodata questions.

Barrett, Phillips, and Alexander (1981) argued that the distinction between predictive and concurrent validity has been exaggerated and that any differences, if present, have a minimal impact on the magnitude of the validity coefficients. Their arguments are supported by a meta-analysis of 366 validity coefficients (Schmitt et al., 1984, p. 412), which suggests that, "contrary to conventional wisdom, the concurrent designs may actually produce

validity coefficients which are slightly superior to predictive designs, especially those predictive designs in which the predictor instruments were used to make hiring decisions."

Unfortunately, the Schmitt et al. comparison of validity designs did not include a separate comparison for the 99 biodata coefficients that were included in their meta-analysis. Consequently, it is impossible to tell from their analysis to what extent the reported advantages of concurrent validity designs apply to biodata, if at all.

There have been few studies that directly compare concurrent and predictive validity designs for biodata. In one early study, Buel (1964) found that predictive validity for biodata was lower than concurrent validity. However, Buel attributed this effect more to intervening variables "sufficient to destroy validity" than to a weakness in the predictive design.

A New York State Department of Civil Service (1987) study may have suffered from similar weaknesses. Concurrent validity coefficients of .28 for performance and .44 for tenure decreased to .18 and .19 in the follow-up predictive study. Although statistically significant, the Department of Civil Service determined that, in this case, the level of validity was insufficient for operationalizing the biodata questionnaire.

The strongest findings come from a recent meta-analysis devoted solely to biodata. Barge and Hough (1988) found that concurrent validity designs consistently demonstrated higher validity than the predictive designs. They found this to be the case for a wide variety of criterion measures.

Mitchell (1990a) suggests that the issue of the merits of concurrent and predictive validity may be more complex than previously thought for biodata. He argues that biodata's scoring strategy in a concurrent design has the potential problem of what he calls "predictive tautology," where predictive tautology refers to biodata's inadvertent use of the same operational variables as both the predictor and the criterion. This tends to inflate the concurrent validity and leads to less stable scoring keys than predictive designs.

Clearly, the results to date are scarce and incomplete. While we know that for cognitive tests, concurrent and predictive validities are very similar (e.g., American Psychological Association, 1987), we need to conduct more research to determine whether these findings for measures of *maximum* performance will hold true for measures of *typical* performance, such as biodata (Schmidt and Rothstein, in press). While the Barge and Hough (1988) study is a useful start, it is important to determine under what conditions each validity strategy is appropriate. More likely than not, each strategy will have its advantages and disadvantages. However, until research suggests otherwise, either approach appears to be acceptable, although the concurrent strategy may be more practical in most instances.

Choosing the Criterion Measure

While much of this paper focuses on the characteristics of biodata that contribute positively to the predictor-criterion relationship, the choice of the criterion is equally fundamental. The term *job success* has been used often in

this paper to refer to the criterion to be predicted. This is because *job success* is a generic term that encompasses virtually any aspect of the job that is considered important by the employer, whether it be performance on critical job tasks as determined by a supervisory rating, production as measured by number of items produced, or employee turnover. Job success is not one thing.

Just as McDaniel, Whetzel, Schmidt, and Maurer (1990) demonstrated that the validity of employee interviews varies as a function of the type of criterion used, the validity of biodata varies with the type of criterion selected (Barge and Hough, 1988). The relevant literature suggests that biodata does not predict all criteria equally well and that biodata may be particularly effective in the prediction of non-cognitive and non-task oriented criteria.

England (1971) lists some of the large variety of criteria that have been successfully predicted by biodata. The list includes: *tenure*, *absenteeism*, and *rate of salary increase*. It also includes many performance production measures, such as *number of sales*, *amount of sales*, *hourly production rate*, and *accident rate*. Supervisory ratings can be even more complex, including not only such measures as *technical knowledge*, *learning ability*, *supervisory ability*, and *overall job performance*, but also such less well-defined measures as *creativity*, *personality*, *responsibility*, and *ability to get along with others*.

The complexity of supervisory ratings was illustrated by Gandy and Mann (1991) in a comparison of supervisory ratings obtained under administrative conditions and under confidential research conditions. Although the biodata questionnaire (i.e., the IAR) could be keyed about equally well to either criterion, the relatively low

correlation (r = .40) between the two criterion measures suggested that they were measuring different aspects of job success. Further analysis suggested that, under administrative conditions, supervisors undervalued academic achievement and overvalued prior experience and that the confidential research ratings were "probably well worth the extra effort" (Gandy et al., in press).

Considerable thought and care needs to go into the choice of any criterion measure. First, the measure needs to be a reliable and valid measure of an important aspect of job success. This is as important as demonstrating the predictor-criterion relationship. Guion (1965) notes that demonstrating that a selection procedure accurately predicts a criterion means nothing, if the selection procedure is predicting the wrong criterion.

Second, any criterion must be fair and not be contaminated or biased with irrelevant variances (Brumback, 1969). Keying biodata items to a biased criterion measure is only likely to perpetuate the bias. Third, Blum and Naylor (1968) add that any criterion of job success should be realistic, relevant, understandable, inexpensive, and acceptable to management.

A criterion measure does not need to represent or encompass *all* aspects of the job domain. On the contrary, the *Uniform Guidelines on Employee Selection Procedures* (1978) and the courts have placed relatively fewer constraints on criterion measures than on predictors (Kleiman and Faley, 1985). Generally, criterion measures are considered acceptable if they are based on a job analysis or if the employer can otherwise demonstrate the relevance of the criterion to the job.

While there are few legal demands on the development of a criterion measure, good practice suggests that the criterion measure be based on a full understanding of the domain of job success and that the relevance of the criterion be clearly demonstrated. A charge that a criterion covers only a limited part of the job, however, could be damaging if subgroups who do badly on this criterion do well on other important parts of the job.

Guion (1987, p. 205) notes that criteria may be global or specific. If one's purpose is "to hire generally good people, a global criterion is useful." "However, if you need to solve a specific problem,...then a more specific criterion is needed. If there is more than one specific problem, then more than one specific criterion is called for. But in most situations, a global measure will serve quite well."

The use of multiple specific criteria is not without potential problems. For example, Hunter and Hunter (1984) cited a study by Tucker, Cline, and Schmitt (1967) that yielded one key to predict supervisory ratings (r = .32) and another to predict tenure (r = .50). Cross-validations of each key on the opposite criterion measure (i.e., the criterion measure they had not been developed from) indicated that each key *negatively* predicted the other criterion measure. Such potential incongruities in biodata keys require that employers be very clear about what it is that they wish to predict.

Part of biodata's appeal is that it theoretically can be made to determine a person's status on virtually any measure of individual variation (Mitchell, 1990a). This includes not only measures of tenure and job performance, but could also include so-called "personality" measures (Mitchell, 1990b), work orientation (Day and Silverman,

1989), employee reliability (Hogan and Hogan, 1989), achievement motivation (Helmreich et al., 1986), or performance on another predictor measure, such as a cognitive ability test (Pannone, 1984).

Given the flexibility of biodata, it is important that employers make clear policy decisions regarding the relative importance of different aspects of job success (Wise et al., 1990). They need to define and decide precisely what it is that they wish to predict. Schultz (1989) suggests that this may not be an easy task, given that available criterion measures usually overlap.

In making policy decisions about selection procedures, employers must consider not only the multiple components of job success that they wish to predict but also the different predictor measures, besides biodata, that are or may already be available. Optimally, Asher and Sciarrino (1974) suggest that selection specialists match point-for-point the individual components of the selection procedure (the predictor domain) with the individual components of the domain of job success. For example, if the ability to learn is an important component of job success, a cognitive written test may be useful in the selection procedure. If interests, motivation to work, or other non-task proficiency criteria are important, a properly developed biodata questionnaire targeted at these components should not be overlooked (McHenry et al., 1990).

Choosing the Validation Sample

To the extent possible, the composition of the validation sample should represent a valid cross-section of the applicants or employees of the job to be predicted. The representativeness of the sample is important for at least two reasons. First, it permits the detection of subgroup differ-

ences and possible biases in the criterion measure before the measure is used to key the biodata questionnaire. Second, it avoids situations where a scoring key is developed on a sample of individuals, whose race and gender compositions may be wholly unlike those of the applicants for the position. For example, a scoring key based on an all-male *incumbent* sample must be examined very carefully to ensure that it does not adversely impact on women in the *applicant* population.

Size of the Validation Sample

A frequent criticism of empirical keying is its requirement for very large sample sizes. This is a legitimate concern and employers are urged not to "cut corners" on the number of individuals used to develop a biodata questionnaire. Depending on the type of biodata questions used, estimates of the number of individuals required to develop a biodata scoring key vary.

England (1971) recommended a minimum sample size of 150, divided into high and low criterion groups of 75 individuals each. Of each group of 75, 50 individuals would be used to develop the scoring key, while the remaining 25 would be used to cross-validate the results. Cascio (1982) similarly suggested sample sizes of about 150 individuals in each group.

Mumford and Owens (1987), however, point out that even though stable percent differences can be obtained with as few as 20 individuals in the high and low criterion groups, operationally, the requirement is for a much higher number. One reason is the need to cross-validate the original key on a separate, independent sample of individuals to verify the validity of the key. Additionally, the samples would need to be large enough to permit

analysis of possible subgroup differences. Moreover, when biodata questions are option-keyed, there would need to be *at least* 20 individuals in the high and low groups *for each response option* of each multiple-reponse biodata item.

These considerations, plus others, such as the need to examine race, gender, and other subgroup differences, typically require larger sample sizes. As a rule of thumb, Mumford and Owens (1987) recommend between 400 and 800 individuals to develop a biodata key. Hunter and Hunter (1984) make a similar estimate (400 to 1,000) and recommend additional large samples every three years to check for decay of the empirical key.

In general, it is advisable to use as many individuals as possible. In order to obtain the large numbers of individuals required, a consortium model seems to be emerging for the development of biodata. In this model, jurisdictions with the same or similar occupations and similar selection needs combine their resources to develop a biodata questionnaire.

Although summing across different jurisdictions and occupations helps achieve the large sample size required for empirical keying and cross-validation, it is an advantage that may not be without cost. The cost for consolidation may be a reduction in the overall predictive validity of the biodata questionnaire. By summing across different elements, the prediction may be reduced to the lowest common denominator. For example, instead of predicting job success in a single job, the biodata questionnaire may predict only those components of job success that are shared among the jobs in the study. On the plus side, summing across heterogeneous elements appears to make keyed biodata instruments more robust and generalizable.

Not all biodata requires the same very large sample sizes. For example, the a priori rational keying approach described by Mitchell (1991) requires sample sizes no larger than those required for the criterion-related validation of other selection procedures. If additional research can confirm and more firmly establish the viability and technology of the rational keying approach, more occupations will become suitable for biodata selection procedures.

In summary, although the required sample sizes for empirically keyed biodata are quite large, the actual sizes required to produce stable biodata depends on many complex factors. Simple adherence to rules of thumb or blind allegiance to any single approach may be inappropriate in some situations. For example, non-operational research on what characteristics of biodata most contribute to biodata's positive outcomes could be accomplished with much smaller sample sizes.

Choosing the Occupation

Most occupations are suitable for biodata prediction, provided that acceptable criteria and adequate sample sizes are available. Just a few of the occupations for which biodata have been used effectively are: police officer, clerical employees, supervisors, managers, welfare workers, sales clerks, production workers, scientists, unskilled laborers, military officers, insurance agents, and real estate agents (England, 1971).

When occupations are grouped, as in the Gandy et al. (1989) study, it becomes important that the measure of job success be relatively generic and applicable to all the jobs in the group. Also, when occupations are grouped, potential biodata items need to be examined carefully for relevance to all the occupations in the group.

CHARACTERISTICS OF BIODATA QUESTIONS

While there are considerable evidences for the potentially positive outcomes of biodata, relatively little is known about what characteristics of biodata contribute to these positive outcomes. Nevertheless, the type of information that is gathered and scored is perhaps the single most important factor in determining biodata's validity, reliability, practicality, job-relatedness, and potential adverse impact (van Rijn, 1980). In this section we will discuss some of the major characteristics of biodata and identify which characteristics seem to contribute most importantly to the development of valid and practical public sector selection procedures.

The characteristics of biodata can be analyzed on a number of levels. Figure 1 showed a categorization of biodata on the basis of the content of the inquiries. Other taxonomies categorize biodata inquiries on various qualitative dimensions or on their physical structure or format.

Researchers have gotten into extensive discussions regarding the characteristics and nature of biodata (Mael, 1992). These discussions often are intended to more clearly differentiate biodata from other selection procedures, background investigations, or from measures of personality, interests, temperament, values, beliefs. Despite these discussions, no clear distinctions or clear-cut taxonomies have yet emerged, particularly not in the public sector.

One of the earliest reviews of the characteristics of biodata is that of Asher (1972). Asher categorized biodata

items along eight qualitative dimensions:

1. Verifiable — unverifiable
2. Historical — futuristic
3. Actual — hypothetical
4. Memory — conjectural
5. Factual — interpretive
6. Specific — general
7. Response — response tendency
8. External — internal.

Asher called biodata questions that generally reflected the first-named characteristics of each dimension *hard items*. Questions more like the second-named characteristic were *soft items*. Based on a review that focused on historical and verifiable biodata between 1960 and 1970, Asher argued that hard items are preferable to soft items, possibly because they reduce the tendencies of respondents to fictionalize their responses and also because hard items may be more representative of the individual.

Despite the endorsement for hard items, Asher noted that Walther (1961), using multiple-choice soft items, obtained remarkably high prediction of the performance and turnover of foreign service clerical applicants. Walther used self-descriptive items related to grades, interests in school, job likes and dislikes (e.g., Which of the following characteristics of a job is *least important* to you?), relationship with parents (e.g., Were your parents: (a) Always very strict ... (e) Never very strict?), social activities, steadiness of employment, and hobbies.

The shortage of information about what characteristics of biodata contribute most to biodata's positive outcomes is particularly acute in the public sector. This is because most

of what is known about the characteristics of biodata is based on inquiries that would not be suitable for use in public sector selection procedures.

Barge (1987) categorized biodata on three dimensions: (1) heterogeneity, (2) behavioral discreteness, and (3) behavioral consistency. *Heterogeneity* refers to the complexity of the question and the extent to which it includes more than one behavioral component. *Behavioral consistency*, not unlike *job-relevance*, refers to the similarity between the behaviors asked about and those required on the job. *Discreteness* refers to whether or not the biodata questions address a single discrete behavior or a more abstract summary characteristic.

In examining the relationship between these characteristics and the predictive validity of individual biodata questions, Barge found that *behavioral consistency* and *behavioral discreteness* contributed substantially to an item's validity. *Heterogeneity,* however, predicted validity in a direction opposite to that hypothesized. Items that were *less* heterogeneous were more likely to produce validity at the item level than *more* heterogeneous items. Barge emphasized, however, that this is not to suggest that heterogeneity of a biodata *instrument* may not still be desirable.

To better suit biodata to the constraints of public sector selection, Gandy et al. (1989) developed five criteria for prescreening biodata questions. Overlapping, but also building on the dimensions identified by Asher, Gandy and his colleagues required that their biodata items be: verifiable, job-relevant, non-intrusive, under the applicant's control, and non-discriminatory.

Although these five criteria were designed to be technically sound, they often derived their importance principally from the legal, social, and practical constraints of the public sector. Once applied to the pool of potential biodata questions (such as those shown in Figure 1), only the inquiries directly focusing on *school and education, work experiences,* and *skills* met all five prescreening criteria, as is shown in Figure 2.

Mael (1992), in an effort to consolidate previous taxonomies and the issues currently being debated, proposed a taxonomy of ten biodata dimensions. The ten dimensions, shown with examples in Figure 3, are:

1. Historical — futuristic or hypothetical
2. External — internal
3. Objective — subjective
4. First-hand — second-hand
5. Discrete — summative
6. Verifiable — non-verifiable
7. Controllable — non-controllable
8. Equal Access — non-equal access
9. Job-relevant — not job-relevant
10. Non-invasive — invasive.

Generally, biodata questions that are characteristic of the first-named attribute on each dimension are more likely to be suitable for use in public employee selection procedures than biodata questions that are more characteristic of the second attribute. In addition to these ten attributes, it is critical that biodata questions be nondiscriminatory and that they do not stereotype applicants on the basis of race, sex, national origin, age, religious or political preference, or handicapping conditions.

Figure 2. Impact of Public Sector Prescreens on the Domain of Traditional Biodata Inquiries

(Adapted from Gandy et al., 1989)

Content Area Taxonomy	Public Sector Prescreens: Behavioral/ Objective	Job-Relevant	Applicant Control	Invasive/ Privacy	Stereotype/ Discrim.
Glennon et al. (1966)					
Classification data	Y	N	N	N	N
Habits and attitudes	N	N	Y	N	Y
Health	Y	Y	Y	N	Y
Human relations	N	Y	Y	N	Y
Money	Y	N	Y	N	N
Parental home, childhood, teens	N	N	N	N	N
Personal attributes	N	N	Y	N	Y
Present home, spouse, children	Y	N	Y	N	N
Hobbies and interests	Y	N	Y	N	N
School and education**	Y	Y	Y	Y	Y
Self impressions	N	N	Y	N	N
Values and opinions	N	N	Y	N	N
	Y	Y	Y	Y	Y
England (1971)					
Personal	Y	N	N	N	N
General background	N	N	N	N	N
Education**	Y	Y	Y	Y	Y
Employment experience**	Y	Y	Y	Y	Y
Skills**	Y	Y	Y	Y	Y
Socioeconomic/Financial status	Y	N	Y	N	Y
Social	Y	N	Y	N	N
Interests	Y	N	Y	N	N
Attitudes	Y	N	Y	N	N
Miscellaneous	N	N	Y	N	N

Note. *N* indicates that the category did not meet the criterion for acceptability. *Y* indicates the criterion was met. The ** indicate categories that met all five criteria for acceptability.

Figure 3. A Taxonomy of Biodata Dimensions

(Adapted from Mael 1992)

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

Mael (1992) includes a detailed discussion of the issues related to each of the ten dimensions. Only some of the major ones will be discussed here. Most importantly, it must be noted that many of the listed dimensions are continua and that many individual biodata questions do not fall on the extreme end points. There are few clear guidelines on what types of items are most appropriate, particularly in the public sector. Depending on the situation, different solutions or combinations of different biodata questions may be called for. Moreover, more than one solution may be acceptable in any particular situation.

Behavioral or Factual Orientation

Generally, biodata focuses on behaviors or facts based on behaviors (e.g., grades or awards received) rather than on thoughts, feelings, values or other less clearly-defined aspects of an applicant. This focus, in part, helps distinguish biodata from most temperament and personality-type measures, which generally probe internal states, feelings, and hypothetical situations.

The psychological literature generally supports the behavioral orientation. Schmitt et al. (1984), for example, found that personality measures were "clearly less valid" than other types of predictors. The behavioral or factual orientation of a predictor measure tends also to more closely match or parallel most measures of job success, that is, *job performances* or *work behaviors*. In addition, inquiries about behaviors and facts are often less threatening and more acceptable to job applicants than vague inquiries about their general internal states, hypothetical situations, or personal feelings (e.g., Johnson et al., 1979; Stone et al., 1989).

Generally, biodata's focus on behavior has been directed towards past behaviors. This orientation relies largely on the well-known axiom in psychology that the best predictor of future behavior is past behavior. There are alternative orientations, however.

Kleiman and Faley (1990) speculated that some of the resistance to biodata may be due to its focus on the past. By focussing on the past, certain individuals may feel unfairly penalized because, either: (a) they did not have the opportunity to exhibit the behavior predictive of later success, or (b) they may have recently changed their behavior patterns. Also, if the lack of past opportunities was due to discrimination, biodata might inadvertently perpetuate the discrimination.

To study the time orientation of biodata, Kleiman and Faley (1990) developed two sets of biodata question: one present-oriented and one past-oriented. For example, while a past-oriented item might be, "Did you often read books for pleasure while you were in school?" a present-oriented item would be, "Do you often read books for pleasure?" In a direct comparison, Kleiman and Faley found that the present-oriented items had validities that were at least comparable, if not superior, to the biodata items that focussed on more distant life experiences. Although this suggests that for certain behaviors, present-oriented items may also be useful, there is little or no evidence for the utility of hypothetical or future-oriented biodata questions.

Wernimont and Campbell (1968), distinguishing between predictors that are *signs* and *samples* of the criterion, suggested that prediction is maximized according to the extent the predictor domain samples the criterion

domain. However, in actual practice, direct sampling is rarely feasible and measurement specialists typically have had to rely on *signs* or behavioral antecedents of the criterion. Biodata is no different from most selection procedures and tends to rely primarily on signs. However, all signs are not equal and the challenge with biodata is to identify those signs that are the best predictors of job success.

It is generally believed that prediction is enhanced according to the extent point-for-point correspondences can be established between the predictor and the criterion. Asher and Sciarrino (1974) argued that the more points there are in common between a predictor and a criterion, the higher the validity. At the extreme, such a model applied to biodata would seek behavioral evidences in a person's past for each of the important behaviors required on the job.

The behavioral orientation of biodata is closely linked to a number of the attributes listed by Mael (1992). Generally, the behavioral orientation is likely to result in biodata items that are historical, external, verifiable, and objective. Behaviors by their nature tend to be more historical, external, verifiable, and objective than most other types of information about the applicant.

Different authors vary in the degree to which they require their biodata questions to assume all of these characteristics. For example, Mael (1992) points out that external events need not necessarily be publicly seen ones or be verifiable. He also notes that while some (e.g., Asher, 1972) prefer verifiable items, Gandy et al., (1989) permit items that are "verifiable *in principle*," such as questions about how applicants thought they would be evaluated by their peers, teachers, and others.

To further complicate the categorization of biodata questions, some biodata questions may be complex, including, more than one component. For example, "My peers would evaluate my ability to lead others as ...?" is a hypothetical question, but the bases for the judgment are historical, observed, external behaviors and events.

Objectivity is also generally favored. Responses merely requiring recall of an event are likely to be more reliable and less subject to self-aggrandizement and other distortions than responses requiring subjective interpretations. Because biodata's information is typically factual and completed in the past, it tends to be verifiable and, consequently, is more likely to be reported accurately (Asher, 1972).

Objectivity also tends to be enhanced when the information solicited concerns only first-hand information the respondent is familiar with and if it requires little or no interpretation. Gandy et al. (1989), however, have not totally excluded interpretive items. For example, asking respondents how they *typically* reacted in certain situations, requires a summary evaluation over a number of real-life events. Although this increases the risk of distortion, it also permits a clearer focus on a single construct or class of behaviors, such as communication skills.

Job-Relevance

From a practical perspective, job-relevance or apparent job-relevance, may be the most critical and controversial requirement in the development of a public sector biodata questionnaire. This is the characteristic that directly determines the acceptability of a biodata questionnaire to applicants and other interested parties.

Although historically biodata questions were selected solely on the basis of whether or not they predicted success on the job, such a strategy alone is not a viable option in the public sector, and the American Psychological Association (1988) and many states (Ash, 1988) have taken positions against the use of biographical data that are not also *logically* job-relevant, in addition to being *empirically* job-relevant.

Logical job-relevance refers to a complex continuum. At one end, it includes the most obvious and transparent linkages between the predictor and the criterion. This is sometimes referred to as *face validity*. At the other extreme, it includes less transparent linkages, possibly involving complex psychological constructs and principles.

Generally, research (e.g., Mitchell and Klimoski, 1982; Quaintance, 1981) has shown that biodata questions formulated on the basis of specific logical hypotheses are more likely to yield significant relationships with the criterion than questions for which there were no such hypotheses. Moreover, Pulakos et al. (1988) recommend that biodata questions be selected specifically for the criterion components that they are designed to predict. For example, biodata questionnaires designed to predict criteria with many cognitive components should focus on biodata that are largely cognitive in nature, e.g., performance in school. Biodata designed to predict tenure, on the other hand, might focus more on an applicant's prior work history or current work habits.

The identification of biographical predictors (signs) or determiners of desired job behaviors is no easy task, and although there is considerable psychological research for guidance, the process often is more an art than a science. Owens (1976) recommends that any questions included on

a biodata questionnaire be based on specific hypotheses about the relationship between the question and the criterion. Measurement specialists must to a large extent rely on their own ingenuity to identify these relationships.

Sometimes, panels of subject matter experts are used to help identify and evaluate the logical relationships between biographical data and desired work behaviors or other measure of job success. This approach has some potential disadvantages. Although Barge (1987) showed that biodata items could be reliably rated on a number of dimensions or characteristics, Crosby and Mitchell (1988) found that job experts were largely ineffective in identifying criterion-valid biodata items. Criterion-valid biodata items that lacked obvious face validity were not usually recognized by the experts, and face valid items identified as valid by the experts "tended to show poor criterion-related validity." Thus, sole reliance on panels of experts may seriously diminish the predictive validity of the selection procedure.

Measurement specialists are faced with a challenge. Although biodata, per se, does not require logical job-relevance in addition to empirical job-relevance, the former is highly recommended, particularly in the public sector. However, logical job-relevance presents a dilemma. First, when biodata questions are selected because their job-relevance is obvious or face valid, Crosby and Mitchell (1988) argue that other more valid questions may be overlooked. Second, many researchers (e.g., Crosby and Mitchell, 1988; Dunnette, McCarthy, Carlson, and Kirshner, 1962; Mumford and Owens, 1987) warn that the most obviously relevant biodata items tend to be particularly susceptible to faking. The more obvious the job-relevance, the more likely applicants are to attempt to enhance their scores by distorting their responses.

Instead of face validity, and in addition to biodata's empirical job-relevance, Mitchell (1986) suggested an understanding-based approach to biodata. This approach involves a trained specialist working with job experts to identify life facts and experiences that can be logically linked to the important behaviors that distinguish successful from unsuccessful job incumbents. The approach is designed to develop items that have clear logical relationships to the criterion, but for which the reasons behind the relationships are not obvious and not intuitive. Nevertheless, the reasoning is such that it could be readily understood by applicants, job experts, administrators, and others, if it were explained.

Mitchell's objective was to avoid sacrificing empirical validity for the sake of face validity. The approach potentially opens the way to a wide variety of biodata items whose content—on the face—appears irrelevant to the target job. The success of the approach, however, will depend on the skills of the analyst and, to a large extent, on the understanding applicants and others have of this approach and their acceptance of its logic and rationale. Mitchell (1991) recently demonstrated that a biodata questionnaire scored according to this approach predicted several measures of job success with an average validity of .29.

Applicant Control and Equal Accessibility

In the interest of fair play, the use of information about a person's past against him or her generally may not be considered acceptable, if the person had little or no control over that past. It is generally not considered fair play to penalize an applicant because his or her father did not finish high school, even if this information is highly

predictive of job success. Consequently, some (e.g., Reilly and Chao, 1982, p. 14) have recommended that items not amenable to change or "improvement" be "eliminated from consideration before empirical keying is done."

There are few or no guidelines to assist the measurement specialist in determining what biographical information is sufficiently under the control of the applicant to be considered a fair question. Stricker (1988) considers it unethical to evaluate on uncontrollable factors. Gandy et al.'s (1989) position is more moderate, and questions about situations for which the opportunities might not always have been equal are not automatically excluded from consideration. It is generally argued that all life events potentially shape later behavior, whether the person deliberately chose the experiences or not. Mael (1992) points out that strict adherence to a *control rule* would lead to the exclusion of most life experiences likely to be related to later behavior, including school and work experiences.

Although the concept of control is ultimately a philosophical question, the amount of control a person had over a past event can be described as ranging from nearly complete control to little or no control. For example, most individuals have little control over the education level of their parents. However, they are much more likely to have had control over the types of elective courses they took in high school.

Certainly, the amount of control that was exercised in any given area depends to a considerable extent on the individual involved as well as the options that were available to him or her. Even the types of courses taken in high school can be significantly limited by the availability of the

courses, the ability to satisfy course prerequisites, etc. Often, other persons exerted the greater control over past life events, such as *number of secondary schools attended* or *sources of educational finances*. Nevertheless, certain life experiences thrust upon an individual may be significant precursors of later success, even if the individual exercised only a small degree of influence (e.g., military service).

While biodata inquiries wholly beyond the control of the applicant are best avoided, not all inquiries need involve equally accessible events. Mael (1992) notes that some consider it unfair to conclude that someone's non-participation in an activity infers that the individual lacks the skills required in the activity. Mael suggests that the experiences captured in biodata *add* something to the person (e.g., new skill, self-confidence). Consequently, it may matter more to be able to identify *who* was enriched by exposure to certain experiences than to determine whether or not everyone had an equal opportunity to have the prior experience. In this sense, biodata is not unlike the information typically solicited on resumes.

Given the subtle complexity of this issue, it is advisable for the measurement specialist to obtain the inputs and views of all relevant "stakeholders" to determine where along the "control" and "access" continua an item must be before it is acceptable for inclusion on a biodata questionnaire. It must also be considered that certain life experiences thrust upon an individual may be significant precursors of later job success. Individuals might receive credit for such experiences, even if they exercised only a small degree of influence.

Considerations for Personal Privacy

Even if valid biodata inquiries are job-related, they may still encounter applicant resistance if the information requested invades the privacy of the applicants or otherwise offends their reasonable sensitivities. The Privacy Act of 1974 and corresponding state laws acknowledge the right to privacy to be a fundamental Constitutional right. Users of biodata are advised to avoid any inquiries that may unduly violate the privacy of any applicant.

There are no clear standards for evaluating the invasiveness of biodata inquiries. Furthermore, privacy rights of individuals are not unqualified under the law. Sometimes, relatively invasive questions may be warranted for certain jobs. For example, questions regarding interpersonal activities may be acceptable for law enforcement occupations but less so for occupations where people work alone.

Most questions pertaining to topics such as race, national origin, gender, religious or political preferences, financial status, age, attitudes towards sex, or marital status would run afoul of a variety of federal, state, and local laws, regulations, or guidelines (Arvey and Faley, 1988; Ash, 1988; and Mitchell, 1989). Generally, these types of inquiries are unacceptable, even though their use in early biodata questionnaires demonstrated that these types of questions often had substantial predictive validity (van Rijn, 1980).

Non-Stereotyping and Non-Discrimination

The use of information that discriminates on the basis of race, creed, color, sex, and national origin are clearly inappropriate or illegal under various federal, state, and local civil rights laws, as well as relevant court decisions. In addition to avoiding discriminatory items, measurement specialists would do well to avoid questions that serve as proxies or indirect indicators of discriminatory information.

Biodata should also be screened to ensure that they do not perpetuate offensive stereotypes about any applicant subgroup. Eliminating offensive stereotypes will not only help minimize potential adverse impact but also enhances the acceptability of the biodata questionnaire.

While potentially discriminatory questions are frequently removed from the pool of possible biodata questions before the keying process, this does not always ensure that all discriminatory items are removed. To further reduce adverse impact, a secondary screening may be desirable to remove remaining keyed items that adversely affect the selection of individuals in different subgroups. While, this secondary screening requires the removal of some valid predictors of the criterion, it is a step that may be well-advised. However, the extent to which it is appropriate to reduce validity to achieve equality of results is unsettled, both legally and philosophically.

As was shown in Figure 2, the five prescreening criteria discussed earlier in this section, severely restrict the content domain of biodata. Nevertheless, several content areas meet the requirements of all five criteria. Figure 4 shows examples of the types of questions Gandy et al. (1989) considered acceptable for use in the public service.

Figure 4. Examples of Biodata Items Suitable for the Public Sector

(From Gandy et al., 1989)

My high school teachers would most likely describe my self discipline as:

a. Superior
b. Above average
c. Average
d. Below average
e. Don't know

The number of high school clubs and organized activities (such as band, sports, newspaper, etc.) in which I participated was:

a. 4 or more
b. 3
c. 2
d. 1
e. Didn't participate

The number of high school courses which I failed was:

a. 5 or more
b. 3 or 4
c. 2
d. 1
e. None

My grade point average in my college major was:

a. I did not go to college or went less than two years.
b. Less than 2.90
c. 2.90 - 3.19
d. 3.20 - 3.49
e. 3.50 or higher

In the past three years, the number of different paying jobs I have held for more than two weeks is:

a. 7 or more
b. 5 - 6
c. 3 - 4
d. 1 - 2
e. None

On my present or most recent job, my supervisor rated me as:

a. Outstanding
b. Above average
c. Average
d. Below average
e. Have never been employed or received no rating.

DESIGNING AND ADMINISTERING A BIODATA QUESTIONNAIRE

Although the technical challenges facing biodata are far from trivial, there are now some techniques that can address and meet many, if not most, of them. These techniques apply largely to the design of the content of the biodata questionnaire. These techniques do not guarantee there will be no legal challenges to the biodata procedure, but they do help reduce their likelihood and, in the event of such a challenge, provide a more reasoned, balanced, and sound defense.

The particular challenge in developing a biodata questionnaire, perhaps more so than for any other selection procedure, is the requirement to balance many considerations. This is complicated by the fact that emphasizing one consideration may come at the expense of another.

Some typical steps in the design of an empirically-keyed biodata questionnaire are the following:

1. Identify a target job (s) suitable for biodata.
2. Conduct a job analysis of the target job(s).
3. Identify or develop meaningful measures of job success.
4. Develop a pool of potential biodata questions hypothesized to be predictive (directly or indirectly) of the measures of job success.
5. Prescreen potential questions for format.
6. Prescreen potential questions for acceptability, job-relevance, non-stereotyping, and other relevant characteristics.

7. Obtain job success and biodata information from a large sample of job *incumbents* (concurrent validity).
8. Identify biodata items and responses that differ entiate successful from unsuccessful incumbents and develop a scoring key. Delete items not predictive of success.
9. Analyze keyed items for fairness and adverse impact and delete objectionable items, as needed.
10. Cross-validate the scoring key on an independent sample of incumbents.
11. Determine how the biodata predictor fits into the entire selection process.
12. Administer the biodata questionnaire to *applicants*.
13. Obtain job success data for all applicants hired and determine the predictive validity of the previously keyed biodata questionnaire.
14. Monitor responses of applicants for effects of social desirability or other distorted response patterns.
15. Revise biodata questions, scoring procedures, and success measures, as needed.

Although this is a highly simplified outline of some of the major steps in the development of an empirically keyed biodata selection procedure, it gives some insight into the complexity and magnitude of such an endeavor. The steps are not necessarily taken in order and each step contains multiple challenges, dilemmas, and competing demands. The process is somewhat simpler for biodata that are

rationally scored or that are designed according to a content validity model of test validation, such as that used in interviews or ratings of training and experience.

Methods of Collecting Biodata

In contrast to background investigations and reference checks where information about applicants is usually obtained orally and from sources other than the applicant (McDaniel, 1989), biodata are typically obtained directly from the job applicant by means of a multiple-choice questionnaire. The questionnaire format is a preferred method of obtaining biodata, because it is particularly efficient and permits the collection of a large amount and variety of information in a relatively short period of time. These features make questionnaires highly desirable, particularly when there are large numbers of applicants to be screened, as is typical for many public sector job openings.

Although biodata questionnaires could be administered in the same way that application blanks are completed by applicants, i.e., at their leisure, this is seldom done. Even though there are few time constraints, biodata questionnaires are typically administered under controlled and standardized conditions, often in conjunction with other standardized selection procedures.

The controlled conditions are intended to accomplish several objectives. First, they permit the administrator an opportunity to introduce biodata, to explain how its works, and to caution applicants against false responses. Second, controlled administrations help ensure that copies of the questionnaire are not compromised. Without this control, unscrupulous applicants might take the questionnaire and

"game" the answers for a later administration. Because alternative and equivalent forms of a biodata questionnaire are seldom feasible, applicants providing false responses reduce the predictive validity of the questionnaire and could possibly give these applicants unfair advantages. Third, it is efficient to administer the biodata questionnaire with other selection procedures. This also gives biodata its appropriate recognition and weight in the selection procedure and helps ensure that all predictive information about applicants will be collected.

The questionnaire format of data collection is not to imply that other methods might not be more appropriate in some situations. For example, if a critical piece of biographical information is not readily recalled by the applicant or if it is likely to be distorted, such information might be derived from an archival source, e.g., a college transcript. Similarly, instead of using a paper-and-pencil medium, biodata inquiries might be obtained via a computer terminal.

Question Format

Earlier we discussed taxonomies of biodata based on the content areas of the questions and the qualitative characteristics of the inquiries. There is another relevant taxonomy that categorizes biodata questions according to their structure or format.

As discussed earlier, biodata questionnaires typically consist of highly structured questions and multiple-choice alternative responses. This format, in large part, distinguishes biodata from traditional job application blanks as well as oral interviews where the inquiries are often unstructured and more open-ended. However, there are many variations on biodata's structure.

Owens (1976) lists seven basic question structures based on the type of response options provided. These seven structures include:

1. Yes — No
2. Continuum, single choice
3. Non-continuum, single choice
4. Non-continuum, multiple choice
5. Continuum, plus escape option
6. Non-continuum, plus escape option
7. Common stem, multiple continua.

Research and experience with different types of questions have indicated that certain types of questions are preferred over others in terms of higher validity, greater stability over time and across applicant groups, and greater adaptability to statistical analyses. In general, the preferred format of biographical questions are questions that are: (1) expressed in multiple-choice format, (2) brief, with clearly numbered options arranged along a continuum, (3) complete, containing all possible alternatives or an escape option, and (4) neutral or positive in tone (Owens, Glennon, and Albright, 1962).

Post hoc content analyses of the items of the IAR with the highest predictive validity provided additional clues about the desirable characteristics of biodata questions (Gandy et al., in press). Of the most predictive items, nearly equal numbers of biodata items were based on self-reports of supervisor judgments, peer judgments, and facts presumed to be documentable.

Sources of Questions

The most ready sources of questions to include in a biodata questionnaire are existing questions, especially questions that have been used and have been proven to be successful predictors of success in similar jobs. There are several large compilations of commonly used biographical questions (e.g., Glennon, Albright, and Owens, 1966). Even though some journal articles and technical reports have provided specific information about questions used in different types of biodata questionnaires, most researchers provide few details about their items or how they were derived. Security reasons and the proprietary nature of many biodata instruments may be primary reasons biodata items are not more widely available.

When pre-existing item sources are not readily available or not appropriate, new questions can be developed directly from the job(s) itself, from literature about the job, or from the relevant psychological literature. Often, questions can be adapted from existing selection procedures. For example, questions asked during an oral interview can often be converted to the multiple choice format of biodata, and Russell and Domm (1990) readily generated 550 items from essays derived from assessment center dimensions. It also seems that *life history* essays and interviews can be relatively quick and inexpensive tools for generating large pools of biodata items with a high probability of criterion-related validity (Russell, Mattson, Devlin, and Atwater, 1990). Also, Mitchell (1986) developed procedures for generating biodata items based on an analysis of the developmental experiences hypothesized to underlie job performance.

Figure 1 provided an overview of some of the types of topics that have been used in traditional biodata measures. While the content and variety of biodata questions appear limitless, the technical, ethical, legal, and practical considerations discussed in this paper place significant constraints on the types of questions suitable for a public sector biodata instrument. Gandy et al. (1989) found that their prescreening process eliminated most of the traditional biodata categories, except for school and educational experiences, work history, skills, and interpersonal relations, as shown in Figure 2.

Whatever the method used to acquire or develop biodata items, it is usually necessary to start with a fairly large pool of biodata questions in the development of any biodata questionnaire. First, many questions will not meet the rational prescreening criteria. Second, the empirical keying strategy may reveal that many questions in the pool do not differentiate successful from unsuccessful employees and therefore would not make useful predictors.

Schmidt, Ones, and Hunter (1992) noted the promise of an alternative strategy for developing a biodata selection procedure. Instead of a full-blown biodata study, Breaugh and Dossett (1989) proposed starting with just a few theoretically relevant and meaningful items. Starting with just four items to predict bank teller turnover, they found that three of the four items were valid: tenure in prior job, similarity of prior job, and graduation from high school.

Expertise Needed

Clearly the development of biodata, as any current selection procedure, is not something to leave in inexperienced hands. Considerable expertise in a variety of areas is typically required.

Ordinarily the expertise required derives from specialized training in psychometric principles, research design, statistics, and tests and measurement. Typically this type of training is found in courses leading to advanced degrees in industrial/organizational psychology. In addition, the expert must have access to or be familiar with the legal and regulatory constraints relevant to the public sector selection system to be designed. Finally, specialized expertise is desirable in the use of biodata for personnel selection, including a full awareness of its advantages and disadvantages.

CONCLUSIONS AND RECOMMENDATIONS

Cronbach (1980, p. 45) wrote:

> *Validation is a process of persuading others that a proposed interpretation is reasonable. Validation is thus forensic as well as investigative. Producing an argument credible to others has always been required of the scientist, but credibility is far more important when community interests are at stake and adversaries can be expected to challenge a claim.*

This advice is particularly relevant for biodata. First of all, biodata looks different from most traditional selection procedures. Although many of the questions pertain to information similar to that obtained during some employee interviews or on application blanks, the highly structured format is unfamiliar to most applicants. Second, the scoring procedure is unfamiliar. Consequently, it is particularly important while introducing biodata into a selection procedure to explain carefully, in lay terms, how it was developed and how it is scored. Dieckhoff (1992) found that introducing biodata as a type of "paper-and-pencil" interview was particularly helpful.

Nevertheless, public employers are advised to ensure that reasonable applicants perceive the biodata questions to be fair, job-relevant, and non-invasive of their personal privacy. This is no short order, given that applicants are likely to vary considerably in their perceptions and tolerances for these issues.

Clearly, care must be taken in developing a biodata questionnaire. An extremely diverse and heterogeneous array of items are liberally referred to as biodata. These

different types of biodata differ dramatically with respect to potential advantages, and many would be clearly inappropriate in a public sector selection procedure. The technical and practical implications associated with the choice of biodata content and the scoring approach selected are far from trivial. Considerable expertise and skill are required to strike an appropriate balance among the requirements for high validity, low adverse impact, acceptability, and resistance to faking. As Davey (1984) pointed out, the issues are not just psychometric and statistical, they are also social.

The potential advantages of biodata are numerous and important. Biodata's potential advantages include, but are not limited to, high and generalizable validities, high reliability, little or no adverse impact, relatively low administrative costs when used with large applicant groups, and the ability to tap into aspects of the job performance domain not usually captured very well by traditional cognitive tests.

Employers in the public sector are encouraged to develop biodata measures, to learn more about biodata's characteristics, and to determine which characteristics will be most useful in different situations. While there is considerable knowledge about the potentials of biodata, much is not directly relevant for the public sector. Gandy et al.'s (1989) research was a major breakthrough, because it demonstrated that a biodata questionnaire—developed within the tight constraints of the public sector environment—could and did yield highly useful and stable validities with relatively little adverse impact.

Greater sharing of information and collaboration among public sector jurisdictions in the development and application of biodata can have substantial payoffs. The

formation of consortia to achieve the large sample sizes needed for empirical keying would be a useful first step. Mael (1992) calls for a clearinghouse of well-documented biodata items. Such documentation might include, in part: (1) descriptions of the characteristics of each item, (2) descriptions of the occupations for which the item was designed and used, (3) descriptions of the criterion measures used, (4) reliability and validity evidences, including cross-validities, and (5) descriptions of the samples used in the development of scoring keys. Dissemination of information about item characteristics could well be done without having to compromise the scoring keys. Much would be gained by such collaborations. "Perhaps the most important gain would be to re-establish the respectability of the much-maligned empirical approach to biodata" (Mael, 1992, p. 787).

In summary, while the biodata technology offers many potential benefits, the development of biodata is becoming increasingly complicated, requiring multidimensional prescreens, alternative scoring systems, and knowledge of constantly evolving social and legal environments. Complex and often ill-defined legal, ethical, and social issues must increasingly be addressed in addition to the technical and psychometric requirements. More than ever, using biodata requires a high level of expertise across scientific, technical, social, and legal domains. Despite these constraints, professionally developed and valid biodata questionnaires can be expected to make positive long-term contributions to the selection of our nation's public servants.

REFERENCES

Aamodt, M.G. and Pierce, W.L. Jr. (1987). Comparison of the rare response and vertical percent methods of scoring the biographical information blank. *Educational and Psychological Measurement, 47* (2), 505-511.

American Psychological Association (1988). Amicus curiae brief for the Supreme Court of the United States: Clara Watson v. Fort Worth Bank and Trust. *American Psychologist, 43* (12), 1019-1028.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Arvey, R.D. and Faley, R.N. (1988). *Fairness in selecting employees* (2nd ed.). Reading, MA: Addison-Wesley.

Ash, P. (1988). *The legality of pre-employment inquiries: A guide to state and other jurisdictional rules and regulations.* Park Ridge, IL: London House.

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

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

Baehr, M. and Williams, G.B. (1967). Underlying dimensions of personal background data and their relationship to occupational classification. *Journal of Applied Psychology, 51*, 481-490.

Barge, B.N. (1987). *Characteristics of biodata items and their relationship to validity.* Paper presented at the 95th Annual Convention of the American Psychological Association, New York, NY.

Barge, B.N. and Hough, L.M. (1988). Utility of biographical data for the prediction of job performance. In Hough, L.M., *Literature review: Utility of temperament, biodata, and interest assessment for predicting job performance*, (ARI Research Note 88-020). Alexandria, VA: U.S. Army Research Institute.

Barrett, G.V., Phillips, J.S., and Alexander, R.A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology, 66* (1), 1-6.

Blum, M.L. and Naylor, J.C. (1968). *Industrial Psychology: Its theoretical and social foundations*, New York: Harper Row.

Booth, R.E., McNally, M.S., and Berry, N.H. (1978). Predicting performance effectiveness in paramedical occupations. *Personnel Psychology, 31*, 581-593.

Breaugh, J.A. and Dossett, D.L. (1989). Rethinking the use of personal history information: The value of theory-based biodata for predicting turnover. *Journal of Business Psychology, 3*, 371-385.

Brown, S.H. (1978). Long-term validity of a personal history item scoring procedure. *Journal of Applied Psychology, 63* (6), 673-676.

Brumback, G.B. (1969). A note on criterion contamination in the validation of biographical data. *Educational and Psychological Measurement, 29* (2), 439-443.

Buel, W.D. (1964). Voluntary female clerical turnover: The concurrent and predictive validity of a weighted application blank. *Journal of Applied Psychology, 48* (3), 180-182.

Buel, W.D., Albright, L.E., and Glennon, J.R. (1966). A note on the generality and cross-validity of personal history for identifying creative research scientists. *Journal of Applied Psychology, 50*, 217-219.

Cascio, W.F. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology, 60* (6), 767-769.

Cascio, W.F. (1976). Turnover, biographical data, and fair employment practice. *Journal of Applied Psychology, 61* (5), 576-580.

Cascio, W.F. (1982). *Applied psychology in personnel management.* Reston, VA: Reston Publishing.

Chaney, F.B. and Owens, W.A. (1964). Life history antecedents of sales, research, and general engineering interest. *Journal of Applied Psychology, 48* (2), 101-105

Childs, A. and Klimoski, R.J. (1986). Successfully predicting career success: An application of the biographical inventory. *Journal of Applied Psychology, 71* (1), 38.

Cohen, J. and Cohen, P. (1975). *Applied multiple regression / correlation analysis for the behavioral sciences.* NJ: Erlbaum.

Colquit, A. and Becker, T. (1989). *Faking of a biodata form in use: A field study.* Paper presented at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Boston, MA.

Corcione, G.K. (1985). *Alternative selection methods: Biodata —A viable option.* Unpublished report, Albany, NY: New York State Department of Civil Service.

Corcione, G.K. and Means, R. (1986). *Biodata research project: The New York State experience.* Paper presented at the 10th Annual Conference of the International Personnel Management Assessment Council, San Francisco, CA.

Cronbach, L.J. (1980). Selection theory for a political world. *Public Personnel Management, 9* (1), 37-50.

Crosby, M.M. and Mitchell, T.W. (1988). The obscurity of biodata predictor-criterion relationships: A blessing in disguise? In T.W. Mitchell (Chair), *Advancing the theory and method of biodata.* Symposium conducted at the 3rd Annual Conference of the Society for Industrial Organizational Psychology, Dallas, TX.

Davey, B.W. (1984) Personnel testing and the search for alternatives. *Public Personnel Management Journal, 13* (4), 361-374.

Day, D.V. and Silverman, S.B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology, 42*, 25-36

Dieckhoff, F. (1987). *The development of a supervisory profile questionnaire.* Paper presented at the 11th Annual Conference of the International Personnel Management Association Assessment Council, Philadelphia, PA.

Dieckhoff, F. (personal communication, February 20, 1992).

Doll, R.E. (1971). Item susceptibility to attempt faking as related to item characteristics and adopted fake set. *Journal of Psychology, 77*, 9-16.

Drakeley, R.J., Herriot, P., and Jones, A. (1988). Biographical data, training success, and turnover. *Journal of Occupational Psychology, 61*, 145-152.

Dunnette, M.D., McCartney, J., Carlson, H.C., and Kirchner, W.K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology, 15*, 13-24.

Dye, D.A. (1991). *Construct validity of a biographical data inventory: A confirmatory factor analysis.* Unpublished doctoral dissertation, George Washington University, Washington, DC.

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

Ferguson, L.W. (1967). Economic maturity. *Personnel Journal, 46*, 22-26.

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

Freeberg, N.E. (1967). The biographical information blank as a predictor of student achievement: A review. *Psychological Reports, 20*, 911-925.

French, J. (personal communication, 1986). In Corcione, G.K. and Means, R., *Biodata research project: The New York State experience.* Paper presented at the 10th Annual Conference of the International Personnel Management Assessment Council, San Francisco, CA.

French, J. (1991). In French, J. (Chair), *Selection methods and prediction of turnover and performance for welfare eligibility workers.* Symposium of the 15th Annual Conference of the International Personnel Management Association Assessment Council, Chicago, IL.

Freudenberg, R. (1991). Whither construct validity? *Newsletter of the Personnel Testing Council of Southern California, 13* (10), 2-4.

Gandy, J.A., Dye, D.A., and MacLane, C.N. (in press). Federal government selection: The Individual Achievement Record. In Stokes, G.S., Mumford, M.D., and Owens, W.A. (Eds.). *The biodata handbook: Theory, research, and applications.* Palo Alto, CA: Consulting Psychologists Press.

Gandy, J.A. and Mann, W.G. (1991). *Perspectives on workforce quality: Administrative and research performance appraisals.* Paper presented at the 15th Annual Conference of the International Personnel Management Association Assessment Council, Chicago, IL.

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

Glennon, J.R., Albright, L.E., and Owens, W.A. (1966). *A catalog of life history items.* Prepared for the Scientific Affairs Committee of the American Psychological Association, Greensboro, NC: Center for Creative Leadership.

Goldsmith, D.B. (1922). The use of the personal history blank as a salesmanship test. *Journal of Applied Psychology, 6,* 149-155.

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

Guion, R.M. (1965). *Personnel testing.* New York: McGraw-Hill.

Guion, R.M. (1987). Changing views of personnel selection research. *Personnel Psychology, 40,* 199-213.

Guion, R.M. and Gibson, W.M. (1988). Personnel selection and placement. In Rosensweig, M.R. and Porter, L.W. (Eds.), *Annual Review of Psychology, 39,* 349-374.

Gulliksen, H. and Wilks, S.S. (1950). Regression tests for several samples. *Psychometrika, 15* (2), 91-114.

Hammer, E.G. and Kleiman, L.S. (1988). Getting to know you. *Personnel Administrator, 33* (5), 86-92.

Helmreich, R.L., Savin, L.L., and Carsrud, A.L. (1986). The honeymoon effect in job performance: Temporal increases in predictive power of achievement motivation. *Journal of Applied Psychology, 71,* 185-188.

Hogan, J. and Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology, 74,* 273-279.

Hough, L.M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69* (1), 135-146.

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., and McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75* (5), 581-595.

Hunter, J.E. and Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96* (1), 72-98.

Hunter, J.E. and Schmidt, F.L. (1976). A critical analysis of the statistical and ethical implications of five dimensions of test fairness. *Psychological Bulletin, 83*, 1053-1071.

International Personnel Management Association (1991, November). Correctional Officer biodata validation project to begin November 1991. *IPMA News*, p. 14.

Johnson, D.A., Newton, N.W., and Peck, L.A. (1979). Predicting tenure of municipal clerical employees: A multiple regression analysis. *Public Personnel Management, 8* (3), 182-190.

Kavanagh, M.J. and York, D.R. (1972). Biographical correlates of middle managers' performance. *Personnel Psychology, 25*, 319-332.

Kleiman, L.S. and Faley, R.H. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. *Personnel Psychology, 38*, 803-833.

Kleiman, L.S. and Faley, R.H. (1990). A comparative analysis of the empirical validity of past- and present-oriented biographical items. *Journal of Business and Psychology, 4* (4), 431-437.

Kluger, A.N., Reilly, R.R., and Russell, C.J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76* (6), 889-896.

Landy, F.J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41* (11), 1183-1192.

Laurent, H. (1962). Early identification of management talent. *Management Record, 24*, 33-38.

Laurent, H. (1970). Cross-cultural cross-validation of empirically validated tests. *Journal of Applied Psychology, 54* (5), 417-423.

Lautenschlager, G. (1985). Within subject measures for the assessment of individual differences in faking. *Educational and Psychological Measurement, 46*, 309-316.

Lee, R. and Booth, J.M. (1974). A utility analysis of a weighted application blank designed to predict turnover for clerical employees. *Journal of Applied Psychology, 59*, 516-518.

Lecznar, W.B. and Dailey, J.T. (1950). Keying biographical inventories in classification test batteries. *American Psychologist, 5*, 279.

Lefkowitz, J. (1972). Differential validity: Ethnic group as a moderator in predicting tenure. *Personnel Psychology, 25*, 223-240.

Lilienthal, R.A. and Pearlman, K. (1983). *The validity of Federal selection tests for aid technicians in the health science, and engineering fields* (OPRD 83-1). Washington, DC: U.S. Office of Personnel Management.

MacLane, C.N. (personal communication, December 12, 1991).

Mael, F.A. (1992). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*, 763-792.

Malone, M.P. (1987). Predictive efficiency and discriminatory impact of verifiable biographical data as a function of data analysis procedure. Unpublished doctoral dissertation. In Mumford, M.D. and Owens, W.A., Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11* (1), 1-31.

Mann, W.G. (1983). *Feasibility of a content valid biographical questionnaire for the selection of municipal police officers.* Paper presented at the 7th Annual Conference of the International Personnel Management Association Assessment Council, Washington, DC.

Matteson, M.T. (1978). An alternative approach to using biographical data for predicting job success. *Journal of Occupational Psychology, 51*, 155-162.

McDaniel, M.A. (1989). Biographical constructs for predicting employee suitability. *Journal of Applied Psychology, 74* (6), 964-970.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L., and Maurer, S. (1990). *The validity of the employment interview.* Paper presented at the 14th Annual Conference of the International Personnel Management Association Assessment Council, San Diego, CA.

McHenry, J.J., Hough, L.M., Toquam, J.L., Hanson, M., and Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

McKillip, R.H. and Clark, C.L. (1974). *Biographical data and job performance* (Technical Memorandum 74-1). Washington, DC: U.S. Civil Service Commission.

Michigan Employment Security Commission (1991). Personal correspondence from Dell Alston, Chief, Northern Test Development Center, September 7.

Mitchell, T.W. (1986). Specialized job analysis for developing rationally-oriented biodata prediction systems. In S. Gael (Chair), *Advances in tailoring job analysis methods for specific applications.* Symposium conducted at the First Annual Conference, Society for Industrial and Organizational Psychology, Chicago, IL.

Mitchell, T.W. (1987). Electronic mechanism for controlling false biodata in computerized selection testing. In L.J. Stricker, L.J. (Chair), *Problems of biodata distortion in personnel selection systems.* Symposium conducted at the 95th Annual Convention of the American Psychological Association, New York, NY.

Mitchell, T.W. (1989). *Are weighted application blanks legal and valid?* Proceeding of the Second Annual National Assessment Conference, Personnel Decisions, Inc. and the Department of Psychology of the University of Minnesota, Minneapolis, MN.

Mitchell, T.W. (1990a). *Developing and using valid and legal biodata selection instruments.* Workshop of the 14th Annual Conference of the International Personnel Management Association Assessment Council (IPMAAC), San Diego, CA.

Mitchell, T.W. (1990b). Can biodata predict personality? In Aamodt, M.G. (Chair), *What does biodata predict?* Symposium of the 14th Annual Conference of the International Personnel Management Association Assessment Council, San Diego, CA.

Mitchell, T.W. (1991). *A priori biodata to predict teamwork, learning rate, and retention.* Unpublished manuscript.

Mitchell, T.W. and Klimoski, R.J. (1982). Is it rational to be empirical? A test of methods for scoring biographical data. *Journal of Applied Psychology, 67* (4), 411-418.

Mobley, W.H. (1982). *Employee turnover: Causes, consequences, and control.* Reading, Mass.: Addison-Wesley.

Moran, G., Walsh, J.A., Clement, W., and Bumbeck, J. (1968). Personal data as a predictor of small finance company credit risk. *Personnel Psychology, 21,* 245-253.

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

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

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

Myers, J.H. and Errett, W. (1959). The problem of preselection in weighted application blank studies. *Journal of Applied Psychology, 43* (2), 94-94

Nevo, B. (1976). Using biographical data to predict the success of men and women in the Army. *Journal of Applied Psychology, 61*, 106-108.

New York State Department of Civil Service (1987). *Evaluating the use of biographical data for screening therapy aide candidates.* Unpublished report, Albany, NY: Author.

Olson, D. M. (1981). *Biographical correlates of a cognitive abilities battery.* Paper presented at the 23rd Annual Conference of the Military Testing Association, Washington, DC.

Orpen, C. (1971). The fakability of the Edwards Personal Preference Schedule in personnel selection. *Personnel Psychology, 24*, 1-4.

Outerbridge, A.N. (1979). *A survey of test validation study costs* (TM-79-18) (NTIS No. PB80-127319). Washington, DC: U.S. Office of Personnel Management.

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

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

Owens, W.A., Glennon, J.R., and 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. and Schoenfeldt, L.F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64*, 569-607.

Pace, L.A. and Schoenfeldt, L.F. (1977). Legal concerns in the use of weighted application blank studies. *Journal of Applied Psychology, 43*, 159-166.

Pannone, R. (1984). Predicting test performance: A content valid approach to screening applicants. *Personnel Psychology, 37*, 507-514.

Pannone, R. (1987). The effects of faking on biodata validity coefficients. In Stricker, L.J. (Chair), *Problems of biodata distortion in personnel selection*. Symposium conducted at the 95th Annual Convention of the American Psychological Association, New York, NY.

Pearlman, K., Schmidt, F.L., and Hunter J.E., (1980). Validity generalization results for tests used to predict training success and job proficiency in clerical occupations. *Journal of Applied Psychology, 65*, 373-406.

Plag, J.A. and Goffman, J.M. (1967). The Armed Forces Qualification Test: Its validity in predicting military effectiveness for Naval enlistees. *Personnel Psychology, 20*, 323-340.

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

Quaintance, M.K. (1981). *Development of a weighted application blank to predict managerial assessment center performance*. Unpublished doctoral dissertation, George Washington University, Washington, DC.

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

Weisen, J.P., Abrams, N.E., and McAttee, S.A. (1990). *Employment testing: A public sector viewpoint.* Personnel Assessment Monograph, 2 (3), Alexandria, VA: International Personnel Management Association Assessment Council.

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

Wise, L.L., McHenry, J., and Campbell, J.P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43*, 355-366