WESER

Personnel Assessment Monographs

■

RECENT INNOVATIONS IN PUBLIC SECTOR ASSESSMENT

■

Volume 2, Number 2
April 1990

ipmaac

RECENT INNOVATIONS IN PUBLIC SECTOR ASSESSMENT

Charles F. Sproule

Editor's Preface

Since its first publication in November, 1986, Personnel Assessment Monographs has become known for its unique combination of professionalism and utility. Each monograph has provided field-tested tools, techniques and solutions, ready for application by the practicing professional. This monograph continues the tradition.

Recent Innovations in Public Sector Assessment was written for submission to the National Commission on Testing and Public Policy (NCOTAPP), which commissioned and provided financial support for the development of the paper. NCOTAPP is expected to incorporate a version of this paper in its final report. Because the report was thoroughly reviewed and revised prior to submission to the Commission, it is being printed with limited editorial revision. For a new editor, this is equivalent to being born with a silver spoon in the mouth. The report should not be interpreted as the official position of IPMA, nor of NCOTAPP. However, readers will notice that the reported innovations come from a broad sampling of IPMAAC member jurisdictions. IPMAAC members also served as reviewers of the report.

The editor wishes to thank NCOTAPP for both its financial support and for providing a forum for the personnel assessment profession. Thanks are also due NCOTAPP for allowing wide distribution of this information to the IPMAAC membership as a monograph. In addition, credit should be given to Cheryle Whisenhunt at the Oklahoma Office of Personnel Management for the time, talent and effort that brought about the camera-ready copy.

It is hoped that readers will use the monograph to discover authors and jurisdictions who have successfully dealt with problems or issues they may now be facing. Readers are encouraged not to overlook the comprehensive list of references. To the extent that the monograph allows jurisdictions to solve existing problems and avoid pitfalls, it has achieved its purpose.

ii

CHARLES F. SPROULE is a personnel assessment manager for the Pennsylvania State Civil Service Commission, where he is Chief of the Division of Test Development and previously was Chief of the Research and Special Projects Division. He served a part-time mobility assignment with the U.S. Civil Service Commission as a Personnel Research Psychologist and has served as a consultant to a variety of federal, state and local agencies.

He holds a masters degree in guidance and counseling, and has completed additional graduate work in research, statistics and testing. Charley has developed, led and instructed a variety of workshops and seminars in assessment.

He is a member of the American Psychological Association, the Society of Industrial and Organizational Psychology, the International Personnel Management Association (IPMA) and the IPMA Assessment Council (IPMAAC). He has been an IPMA Chapter President, a founder and President of IPMAAC, and the first President of the Mid-Atlantic Personnel Assessment Consortium (MAPAC). He is a recipient of IPMAAC's Stephen E. Bemis award for significant practical contributions to the personnel field.

Personnel Assessment Monograph

RECENT INNOVATIONS IN PUBLIC SECTOR ASSESSMENT

CONTENTS

INTRODUCTION AND EXECUTIVE SUMMARY

This introduction and executive summary describes why and how this paper was prepared, provides background information on public sector assessment innovations, summarizes and comments on the findings of this report, and summarizes the recommendations found in the various sections of this report.

INTRODUCTION

The National Commission on Testing and Public Policy is a blue-ribbon body consisting of leaders in education, training, human resource development, public and private sector employment, military personnel policy, government, and law. The Commission is conducting a study of the role of testing in the allocation of educational, training, and employment opportunities. The major outcome of the Commission's work will be a report that focuses on the social, economic, and political contexts, uses, and consequences of testing. To develop the knowledge base for its report, the Commission has invited papers on a variety of testing issues; and sponsored hearings, seminars and meetings to elicit information from as many testing publics as possible.

In December 1988, the National Commission on Testing and Public Policy invited the International Personnel Management Association Assessment Council (IPMAAC) to author two papers of interest to the Commission. In January 1989, Charles F. Sproule, a past-president of IPMAAC and Chief of Test Development for the Pennsylvania State Civil Service Commission, accepted the request of Dr. Joel P. Wiesen, 1988-1989 IPMAAC President, to undertake the preparation of this paper on recent innovations in public sector assessment. An additional IPMAAC paper on "Employment Testing: A Public Sector Viewpoint," which was prepared by the current IPMAAC President, the Past-President and the President-elect, will also be available.

The International Personnel Management Association (IPMA) is a nonprofit, educational organization representing

1

and serving the interests of personnel professionals in the public sector. There are 55,000 human resource professionals who are either individual members of the Association, or employed by one of IPMA's agency members.

The International Personnel Management Association Assessment Council (IPMAAC), formed in 1976, is a professional section of IPMA. IPMAAC has approximately 560 members. The majority of IPMAAC members work in professional level personnel assessment in federal, state, city, and county governments. Most IPMAAC members work for governmental central personnel agencies or personnel offices and have responsibility for such functions as employment testing, promotional and performance evaluation programs, organizational development, and other assessment functions. Some members work in the private sector, and in colleges and universities.

The purposes of IPMAAC, as stated in the bylaws of the organization, include:

- to encourage and give direction to public personnel assessment improvement efforts;
- to encourage and facilitate intergovernmental cooperation, information exchange, and resource sharing; and
- to contribute to the formation of public policy relating to public personnel assessment (IPMAAC Bylaws, 1981, p. 1).

The public sector and the National Commission on Testing and Public Policy can benefit from contributions of the assessment profession which are illustrated in the work described in this paper, and in the IPMAAC paper on "Employment Testing: A Public Sector Viewpoint."

The information for this paper was collected in a variety of ways. In December 1988, a request for information on public sector assessment innovations was sent to all IPMAAC members with the Winter 1988 issue of *Assessment Council News*. In addition, a letter and a questionnaire were sent to 47 IPMAAC members requesting information on innovations. Those who were sent the letter and questionnaire included: all IPMAAC officers and committee chairs, presidents of all regional public sector consortia,

presidents of related assessment organizations with large public sector representation (such as personnel testing councils), and selected IPMAAC members from federal, state, city, and county governmental agencies known to be at the leading edge of public sector assessment. A review of proceedings of annual IPMAAC conferences since 1980, IPMAAC newsletters, other public sector literature, and selected assessment-related literature was also conducted.

This paper reports primarily on merit system innovations in public sector assessment. It does not report in detail on other public sector assessment efforts, such as those in public education, the military, or the Job Service of the U.S. Department of Labor.

The public sector has been at the forefront of developments in employment testing and other areas of personnel assessment. For example, the U.S. Civil Service Commission, now the Office of Personnel Management (OPM), issued the first major publications on such recent developments as meta-analysis, validity generalization, the utility of alternative selection procedures, and the behavioral consistency model of rating training and experience. The public sector has led the exploration in the search for alternative selection procedures which have less score differences for minorities than traditional paper and pencil tests. Many of the assessment procedures described in this paper meet that goal.

The public sector has also been the prime supporter of job-specific tests based on detailed job analysis and rigorous content validity based test development procedures, such as those described in the IPMA publication "Content Validity: A Procedural Manual" (Mussio & Smith). In addition, a wealth of criterion related research studies have taken place in the public sector. For example, most research on the various methods of rating training and experience has been conducted in the public sector.

"Personnel selection for public organizations typically entails greater complexity than selection within the private sector" (Barrett et al., 1978). Employment procedures and standards in the public sector are regulated in more detail and are more open to close review and public scrutiny.

3

Public sector employment procedures must usually meet merit system requirements. This has led to developments in areas of personnel assessment which are somewhat unique to the public sector. Examples of this can be found in the articles in this paper on "Minimum Qualifications," "Structured Oral Examinations," "Ratings of Experience and Training," "Biodata," "Use of Test Scores," and "Legal Provisions to Encourage Innovations and Research..."

Because of the limited time available to obtain information and prepare this paper, many noteworthy innovations could not be included. Examples include: a project in San Bernadino County, California, on the development and application of job-specific medical standards; a variety of efforts to improve scaling and rating procedures, and procedures for establishing passing scores; cooperative multi-jurisdictional efforts to improve assessment through such activities as cooperative test development, test exchange and test transportability, conduct of training for assessment specialists, a national job analysis of assessment specialist positions; research on test item bias and successful efforts to reduce group differences on written tests while maintaining validity; large scale adverse impact analysis studies and what we have learned from them; use of job knowledge tests in place of ability tests to reduce score differences between groups; and many other public sector activities and research related to the interests of IPMAAC and the interests of the National Commission on Testing and Public Policy.

Two of the National Commission on Testing and Public Policy reviewers of this paper noted that paper and pencil written tests were not included. For example, Kaye Evleth stated: "...the multiple choice written test should not be discarded as a viable selection method when it is used in a responsible manner" (Evleth, correspondence, 5/19/89). Dr. Lila J. Quero noted that the development of job-related paper and pencil examinations is the most attainable assessment alternative for 90% of public jurisdictions because of the higher cost and expertise required for development and use of many alternative testing methods.

4

Since this paper has the objective of presenting information on innovations, the traditional paper and pencil testing mode was not included. However, there is strong evidence for the validity of job-specific paper and pencil job knowledge tests. Such tests were found to have high validities in a recent meta-analysis. "Validity is nearly twice as high for job-specific tests than for off-the-shelf tests. ...Implications for selection would suggest to employers that when tests of job knowledge are used, there is much to be gained by developing them to be job specific" (Dye, Reck, & McDaniel, 1989, p. 9). Job knowledge tests with high job-test similarity were found to have a corrected mean validity of .62 (.31 uncorrected, 59 correlations, 3,965 subjects) for predicting job performance, and .76 (.49 uncorrected, 13 correlations, 1,525 subjects) for predicting training success (Dye, Reck, & McDaniel, 1989, p. 12).

Another testing mode not reviewed in this paper because of common use in the public sector is physical work sample tests.

Readers of this paper need to keep in mind that there is "...an extensive array of testing methods.... There is not one method that is correct for every job, nor is one of the methods correct for all jobs." (Evleth, correspondence, 5/19/89) No one method is the answer to all selection problems.

SUMMARY OF FINDINGS

Following is summary information on the findings presented in each section of this paper. The summary information is followed by comments on the challenges, trends and research findings reported in this paper.

Selected Assessment Methods

The initial sections of this report review selected assessment methods. The reviews illustrate the wide variety of procedures used in the public service and the evolution of methods and knowledge which has occurred in the past decade. Each section contains a variety of examples describing public sector methods, as well as a summary of research findings. Each section ends with a brief summary and contains recommendations relating to the topic covered for the consideration of The National Commission on Testing and Public Policy.

Minimum Qualifications (MQs). MQs are the typical first screen or "test" in the employment process. There has been a history of litigation which demonstrates that MQs are often set at levels beyond what is justified. This results in unfair discrimination.

Public sector jurisdictions have applied content validity procedures to the development of MQs to help assure job-relatedness. Qualifications developed following these procedures are more likely to be fair to job applicants than traditional procedures for developing minimum qualifications requirements. Training courses have been developed by public sector assessment specialists in the technology of proper MQ development.

Guides and procedures have also been developed for reviewing and evaluating applications and resumes in a manner which considers alternative ways for candidates to demonstrate possession of minimum requirements. Considerable flexibility exists in applying MQ requirements in the public sector. An example of such a program in the state of Pennsylvania is described. The state of Delaware's knowledge, skill and ability approach to establishing MQs is a very recent innovation described in the MQ section of the

7

paper. The Delaware procedure expands on a previous federal procedure developed for blue collar jobs.

Rating of Training and Experience (E&T). This examination procedure is covered next. In the private sector heavy reliance is placed on the review of applications and resumes. This process has been quantified and researched in the public sector. Research results are moving public jurisdictions away from practices of only a decade ago. The research has identified methods, such as behavioral consistency and self-ratings, which have higher levels of validity than traditional methods and which appear to be less likely to screen out minorities. This section of the paper presents examples of public sector use of and research on various E&T methods. The heavy reliance on application and resume information in the private sector is more likely to adversely affect minority groups than the innovative methods developed in the public sector.

Biodata. This selection procedure has been historically developed based on a strictly empirical methodology. Use of biodata instruments has been common in the private sector but infrequent in the public sector. These instruments are economical to administer to large applicant groups and easy to machine score. Research results show high levels of validity for a wide variety of occupations and criteria of job success, and low or no adverse impact for biodata instruments.

A few examples of public sector use of and research on biodata instruments are presented in the article. The examples are from federal, state and local governments. A consortia effort is also described. The limited public sector use of biodata is due to such problems as: feasibility, since large samples are needed and extensive developmental and follow-up research is required; lack of job-relatedness, since most instruments have not been developed based on job analysis; possible fakability problems; and possible invasion of privacy problems, since items are often based on personal, social or economic considerations.

Recent research on biodata reported on in this paper has resulted in the biodata procedure being refined and embraced by the federal government for entry-level posi-

8

tions at the professional level. The research has addressed many of the problems outlined above.

Structured Oral Examinations. This procedure has been used extensively in federal, state and local government merit systems for decades. It is used for a wide variety of jobs and job requirements and has high applicant acceptance. Structured oral examinations can assess many important job requirements which cannot readily be assessed by paper and pencil tests. These procedures have recently been newly "discovered" by many measurement professionals.

A content validity model that establishes job-relatedness evidence is typically used to develop structured orals in the public sector. Public sector publications describing this methodology are referenced in this paper.

There has been extensive public sector research on structured oral examinations. The research is summarized in this paper. Examples of public sector use of structured oral examinations are also presented. The research shows that well-developed and carefully administered structured oral examinations using job-related questions, well developed rating scales, and well trained raters have a level of validity comparable to that of cognitive ability tests, and have low or no group differences. However, the procedure is administratively time and resource consuming compared to paper and pencil tests, especially for large applicant groups.

Work Simulation, Work Sample and Performance Tests. These are increasingly common procedures in the public service. A number of public sector guides for the development and conduct of such tests have been prepared and are referenced in this report. The section contains examples of the wide range of occupations and types of job requirements for which the methods have been effectively used. Readers will find many innovative tests in the examples presented.

These methods have high applicant acceptance, high job-relatedness and validity, and typically result in low or no group differences when compared to paper and pencil tests. The major disadvantages of these assessment procedures are the developmental effort required, and the administrative time and costs associated with conducting and scoring

them. Examples of innovations which increase the efficiency
of these methods are reported in this paper including:
cooperative consortia efforts, machine scoreable simulations,
video taped tests, mini-assessment centers, latent image
simulations, "take home" tests, and cost analyses to deter-
mine the break-even point for use of work sample tests as
compared to other procedures.

Selected Federal Assessment Innovations

This article describes a variety of U.S. Office of Person-
nel Management initiatives including: revisions in entry-
level testing procedures, test simplification, use of logic-
based measurement, and a work force quality assessment
research program. These efforts are aimed at such goals as
increasing test quality, speeding up and simplifying the as-
sessment process, and increasing the diversity of the federal
work force. Other federal innovations are described
throughout this report.

The article on federal innovations also summarizes some
critical reviews of federal hiring and assessment practices,
and provides an innovation identified by the U.S. Merit Sys-
tem Protection Board. The section also describes actions
taken by the U.S. Office of Personnel Management related
to the problems identified.

Some of the problems of assessment professionals in the
public sector are noted, including: the reduction of resources
devoted to assessment improvement efforts, such as funding
reductions under the Intergovernmental Act and a weaken-
ing of OPM through the reduction of staff and functions;
negativism towards government employees and government
employment; a decrease in the helping role of OPM for state
and local governments; and the lack of consistent federal
policy direction. Despite the many problems identified,
OPM continues to be a primary contributor to public sector
assessment innovation. Many examples of OPM contribu-
tions, innovations, and research are cited throughout this
report.

Application of Technology to Assessment

Two articles review the application of technology to public sector employment testing. The "Use of Video Technology" and the "Use of Computer Technology" in assessment illustrate creative ways to improve assessment accuracy, and increase job-relatedness and efficiency. A variety of federal and state tests for various occupations and types of job requirements are presented as examples.

Video and computer technology is being used for realistic job previews and work simulation testing, as well as in other areas of assessment such as candidate orientation and training of examiners. Expanded application of these technologies in employment testing can result in more realistic assessment exercises and increased standardization of the tasks presented to examinees. These technologies can reduce costs when work simulation tests are administered to large groups. For example, scoring of video taped candidate responses can increase the efficiency of use of raters.

Computers facilitate the work of assessment professionals in such areas as: word processing, applicant tracking, job and test analysis, item banking, improving test quality, training, personnel management information systems, candidate and employee guidance systems, and providing rapid service to job seekers and employing agencies. Examples of the application of computer technology to these areas are described. However, the application of video and computer technology to assessment is in its infancy in the public sector. Recent advances in technology, such as computer interactive video, the increased speed and storage capacity of microcomputers, lower hardware and software prices, and the availability of more sophisticated software, can lead to rapid advances in the application of these technologies in the public sector.

A number of public sector studies have shown high applicant acceptance and low adverse impact for computer and video based assessment procedures.

Use of Test Scores

Merit system use of test scores has changed dramatically in recent years. The use of tests on an absolute ranking

basis, with little room for judgment, is no longer the norm in some public sector jurisdictions. This trend is evident both in law and regulation and in public sector practice. However, some public sector jurisdictions still are bound by regulations which require rigid reliance on test scores.

Guidelines for the proper use of test scores have been published in a special issue of *Public Personnel Management* on "Assessment Techniques and Challenges" (1984, Volume 13, No. 4). These guidelines generally advocate the use of test scores for personnel selection on a category or banding basis rather than an absolute ranking basis. The guidelines can help employers decide on the most appropriate method of use of test scores.

Legal Provisions to Encourage Innovations and Research and Demonstration Projects

The federal government and some state and local governments have enacted laws and issued administrative procedures to allow waiver of employment and training regulations for the purpose of research, demonstration, and innovation to improve selection and other personnel procedures. Examples of such laws and regulations, and projects conducted under them, are described in this paper. Many such projects resulted in significant improvements in public personnel assessment.

Unfortunately, federal funding support for innovations and research in public sector assessment has been drastically reduced in the federal service. Federal funding of programs to assist state and local governments under the Intergovernmental Personnel Act has been essentially eliminated despite the past success of this program and the provisions of the law which provide for such assistance.

Employment Testing of Persons with Disabilities, and Employment Programs for the Disabled

A variety of efforts are being made in the public sector to modify testing procedures so that they fairly assess special candidate groups. The International Personnel Management Association Assessment Council has published a monograph, described in this article, which include "Model

12

Guidelines for Accommodated Testing of the Disabled." The monograph also describes testing programs for persons with disabilities in some states and in federal governments of the U.S. and Canada. The second section of that chapter describes public sector jurisdiction policy decision-based programs to provide jobs to applicants with special needs, and also provide training and other support. Examples of jurisdictions which carry out both accommodated testing programs and employment programs for the disabled are noted in this article.

Comments on Findings

The article on the disabled, and the section of the Experience and Training Rating paper on grade point average use, illustrate one of the most difficult challenges assessment professionals face. Organizational goals and values are often multiple. Maximizing organizational efficiency through the employment of persons most likely to be high performers is one goal of organizations and assessment professionals. This goal is illustrated by the considerable public sector assessment work in the area of utility research and the research to compare the relative validity of alternative assessment methods. Another goal is to recognize the value of diversity in society and use assessment procedures which provide opportunities for all groups in society to participate fully in the governmental work force. This goal is illustrated in public sector efforts to use assessment procedures which allow for diverse responses in efforts to minimize test score differences between majority and minority groups and in efforts to train and employ members of underrepresented groups.

Recent research results, as summarized in this paper, demonstrate that we have identified assessment methods which can contribute significantly to accomplishing both of the goals mentioned above. Assessment procedures exist which can meet both goals. However, in general, increasing the fairness and flexibility of assessment methods, like any improvement, often results in increased costs. Procedures with high validity and low or no adverse impact are usually more expensive to develop and administer, and more time

consuming to carry out than traditional testing methods. Public policy makers need to commit the resources which will allow increased use of procedures which can help to accomplish both goals.

Dr. Lila J. Quero, who reviewed the first draft of this paper for the National Commission on Testing and Public Policy, estimated that the costs of development and the expertise needed for the use of many of the alternatives reviewed in this paper are beyond the resources of 90% of public jurisdictions (L. Quero, correspondence, 5/89). In the judgment of the author, this is an underestimate of the creativity and proficiency level of public sector jurisdictions. As noted throughout this report, there are a number of jurisdictions using the procedures reviewed. However, additional resources are needed for increased development and broader application of innovations in the public sector. Dr. Quero noted that "...support for increased resources is needed if improvements in testing are to come through." In addition, the innovations need to be publicized to make more employers aware of innovative methods and to encourage broader application.

One of the IPMAAC reviewers of this paper, Dr. Barbara Showers, hypothesized that the reason behavioral consistency E&Ts and structured oral examinations reduce adverse impact is because they allow more diverse responses to be correct. This hypothesis may also apply to work simulation and some of the other testing methods reviewed in this paper. Another contributing factor may be the training provided to raters in how to recognize and avoid bias, and efforts taken to obtain representation on rating boards.

A continuing challenge to the assessment professional will be to find cost-effective ways which allow use of the methods which have been identified as having high levels of validity and low or no adverse impact. This challenge will be with us for some time because of expected continued shortages of public sector funding, given the federal deficit situation, and the *Work Force 2000* projection of a doubling of the percentage of minority group members in new entrants to the labor force.

14

Some agencies and jurisdictions have developed policies and procedures to more directly accomplish the goal of a balanced work force. For example, the "within-group" scoring procedure of the U.S. Employment Service and the "3 + 3" rule of the State of Washington result in separate lists of candidates by majority and minority groups and use top down hiring within each group. Procedures such as these represent public policy decisions which go beyond the authority of public sector assessment professionals. Such procedures attempt to balance the goal of prediction of job performance and the goal of work force representation while minimizing possible loss in utility. These procedures, however, may be subject to legal challenges given recent U.S. Supreme Court decisions.

There are differences in the direction of some trends and innovations identified in this report. One trend is the development and use of tests which are very job-specific, such as: work simulations, behavioral consistency E&Ts, and structured situational oral examinations. Another trend is to group job classes together into broad families and use ability tests which are "tailored" to some degree to an occupational area. The latter trend has been characterized by some users as a more efficient and cost-effective method. These users cite meta-analysis findings (Schmidt & Hunter, 1981) of high validity for ability tests for all jobs in all settings, with evidence of fairness to minority groups, and evidence that use of ability tests can result in large cost savings. Recent data on group differences by assessment method indicate that more job specific tests, such as work simulations and structured job-related orals, generally have lower group differences in scores than ability tests, and have a similar or perhaps higher level of validity.

The above two trends also represent a different emphasis in terms of reliance on the type(s) of validity evidence necessary to support the assessment procedure. The ability measures may require criterion-related or validity generalization evidence; content evidence may suffice for the more job-specific simulations. It will be interesting to see which of the two trends represent where public sector assessment will be in the next century. It is likely that the ability

15

approach will be dominant for some employers, unless increased resources are made available or policy directions change.

SUMMARY OF RECOMMENDATIONS

Following is a list of the recommendations contained in this report. The reasons for these recommendations, and the data which support them, can be found in the various sections of this report.

The specific recommendations listed below are presented not in the same order as they appear in this paper, but rather are grouped under three broad recommendations.

1. The National Commission on Testing and Public Policy should recommend public policy to encourage the expansion of innovation, research, and diversity in employment testing improvement efforts. Increased funding and organizational support should be provided for such efforts. Specific recommendations include:

   — A clear consistent policy direction for public employment assessment. For example, a clear consistent policy direction on use of tests in the federal government and use of tests in merit system assessment is needed.

   — Increased intergovernmental cooperation, such as consortia projects and research, to reduce the developmental costs of innovative assessment procedures for individual employers, and to expand use of innovative procedures.

   — Expansion of the federal, state, and local resources being applied to assessment improvement efforts.

   Federal funding to support and encourage research and intergovernmental cooperation, such as the examples cited in this paper, should be restored to the levels provided in the 1970s through the U.S. Civil Service Commission, now the U.S. Office of Personnel Management (OPM).

The state and local section of OPM as well as OPM regional and central office assessment and research staff and functions (e.g., the Office of Personnel Research and Development, and the Bureau of Intergovernmental Personnel Programs) should be restored to the levels of complement, funding, and research of the 1970s. These resources provided the basis for many current innovations. This includes the restoration of Intergovernmental Personnel Act funding for assessment improvement efforts, technical assistance, and training to federal agencies and state and local governments.

- The adoption of legislative and administrative provisions which allow for innovation such as the waiver of merit hiring requirements for research and demonstration purposes.

- The adoption of guidelines on test modification and test accommodation for the disabled by employers and educational institutions. The Commission should also encourage public policy makers to more directly address the needs of particular groups through special employment and training programs.

2. The National Commission on Testing and Public Policy should recommend increased use of and research on assessment methods which have been identified as having high levels of validity and low or no adverse impact. Specific recommendations include:

- Use of content validity-based procedures in determining minimum qualification requirements. Such procedures can increase the appropriateness and fairness of job requirements, and reduce adverse impact. This recommendation can be followed without large expenditures by employers.

- Increased use of Rating of Experience and Training (E&T) methods which have been shown to have higher levels of validity and lower levels of adverse impact. The behavioral consistency method should be strongly recommended for jobs where it is appropriate. Use of typical application and resume in-

17

formation, such as is usually done in industry and in the traditional merit system point methods of rating training and experience, should be discouraged.

— Continued research on biodata instruments, given the high levels of validity and low levels of adverse impact of well researched and carefully developed instruments. Consortia activities to pool resources in order to conduct needed research should be encouraged.

— Increased use of the structured oral examination methods developed and researched in the public sector. Such procedures are a vast improvement over typical employment interview procedures and are likely to help reduce adverse impact.

— Increased development and use of work sample tests. Innovations and research in this area of assessment should be encouraged.

— Expanded support for and encouragement of the application of video and computer technology to assessment. Continued innovation in the use of these technologies. Research on the cost-effectiveness, relative validity, and adverse impact of tests using these technologies should be encouraged.

3. The National Commission on Testing and Public Policy should recommend the appropriate use of tests and test scores. Use of well developed tests should be encouraged. Tests can be helpful aids in the decision-making process and can contribute greatly to accurate prediction.

Rigid reliance on test scores only in making employment decisions should be discouraged. Public sector agencies which have rigid laws or regulations on the use of scores, such as rules of one or rules of three, should be encouraged to modify such provisions. The National Commission on Testing and Public Policy should recommend the development of public policy on the use of test scores.

SELECTED ASSESSMENT METHODS

Minimum Qualifications (MQs)

Background Information

When applicants seek jobs, the first hurdle they typically face is the employer's minimum training and experience requirements, or other forms of minimum qualifications (MQs). Such requirements are, in effect, pass/fail tests. Applicants who do not meet the minimum qualification requirements are eliminated from further consideration.

Minimum qualifications may include such requirements as:

- Education (e.g., a high school diploma or equivalent, Bachelor's Degree in Accounting)
- Experience (e.g., two years of clerical experience, one year as a supervisor of payroll and benefits)
- Licensure, registration or certification

As part of the applicant evaluation and screening process, minimum qualification requirements must meet legal and professional standards. Qualification requirements are covered under a variety of laws and regulations, including Title VII of the Civil Rights Act and the Uniform Guidelines on Employee Selection Procedures.

Accurate, fair, job-related and valid minimum requirements are critical to assuring equal employment opportunity. There has been a history of litigation which demonstrates that inappropriately high minimum qualification requirements are sometimes used by employers (Amick, 1980).

Employers typically attempt to defend challenged MQs on the basis of content validity, rational arguments, and business necessity arguments. Public safety requirements and legal requirements also sometimes form the justification for MQs.

It has been difficult for employers to successfully defend absolute education and experience requirements, since, as stated in the Griggs case by the Supreme Court, "History is filled with examples of men and women who rendered highly effective performance without the conventional badges of accomplishment in terms of certificates, diplomas, or degrees." Furthermore, most employers have not had a technology or documented procedure to determine what an appropriate job-related minimum qualification should be for a class of similar positions.

Also, there is little empirical data available on the relationship of minimum qualifications requirements to job performance. This is because it is usually not technically feasible to conduct research on this relationship. Applicants who have training and experience below the minimum qualification are usually not given the opportunity to demonstrate if they can adequately perform the job.

Comprehensive reviews of the literature and meta-analysis studies, such as those summarized in the next section of this paper on ratings of experience and training, have examined the relationships of education and experience to job performance and training success. Such reviews have not examined the question of the empirical validity of MQs. Studies have generally found low to, at best, moderate positive relationships between measures of education and experience, and job performance. For example, a recent literature review suggested a low relationship between measures of education and job performance. The same study contained a new meta-analysis of education and job performance relationships which suggests that levels of validity for measures of education are essentially the same or higher than that suggested by the literature review (Dye & Reck, 1988, abstract). Traditional point methods of rating training and experience rely heavily on crediting types and years of education and experience. The criterion related validity of the traditional point method is low (McDaniel & Schmidt, 1985, abstract). On the relationship of experience to job performance, "There is little empirical evidence to provide guidance, and content relevance of prior experience is thus critical" (Abrams, Ash & Johnson, 1986, pp. III-14). Ex-

perience <u>may</u> reach useful levels of criterion related validity in applicant groups which have low mean experience levels (McDaniel, Schmidt & Hunter, 1988, p. 305). Again, however, available data does not provide adequate information on the criterion related validity of absolute minimum qualification requirements.

A common practice in establishing requirements has been to obtain the opinions of supervisors and managers on what the minimum qualification standards should be. Usually, their answers reflect their own training and experience, or the training and experience of those previously hired. Such a method of establishing requirements is likely to result in requirements which are narrow and which perpetuate the status quo. Such requirements are not likely to be inclusive of all possible valid job requirements based on what is necessary to do the work.

In some cases employers, particularly those who have structured job classification and pay systems, use minimum qualification requirements as one basis for determining the pay rate of a job. This often leads to inflating the minimum qualification requirements to obtain a higher pay rate for the job. Such practices may explain why minimum qualification requirements for some jobs have been found not defensible when challenged.

Innovations

Within the past decade, many public sector employers have applied content validation procedures to the development of minimum qualification requirements.

In essence, such procedures start with a job analysis of the duties and tasks of the positions. Next, required knowledges, skills and abilities (KSAs) are derived from the job tasks. Job experts participate in this process to assure accuracy. The knowledge, skill and ability requirements are then evaluated to determine which of them can appropriately be used as the basis for minimum qualifications requirements. For example, only important or critical requirements which are needed at entry should be used as the basis for minimum qualifications standards. Ways in which applicants can demonstrate possession of the entry-level, im-

21

portant KSAs which are to be assessed through the minimum qualifications standards, are defined and used as the basis for determining minimum requirements.

It is not necessary to use all job requirements determined by the job analysis in establishing minimum qualifications. Some requirements may be assessed by the minimum qualifications standards; other requirements may be assessed in other ways, such as in the employment interview.

Following are examples of public sector training programs, and materials on job-content based procedures for establishing job-related minimum qualification requirements. These materials more fully elaborate and explain the procedure outlined above.

Training Programs

Detailed guidance on the various steps in the process of developing job-related minimum qualification standards is provided in a recent International Personnel Management Association Assessment Council training program (Abrams, Ash, Johnson, 1986). The minimum qualifications section of this seminar covers such topics as:

— summary of research evidence on the relationship of educational level, years of education, and years of experience to job performance;

— use of job analysis procedures for developing job-related minimum qualification requirements;

— determining when a degree or diploma requirement is appropriate;

— writing the experience requirement in terms of work performed rather than job titles;

— providing optional ways for candidates to demonstrate possession of the requirements;

— and use of equivalencies.

This IPMAAC seminar has been offered throughout the U.S., including Hawaii, over the past three years.

A pre-conference workshop on "How to Develop Job-Related Minimum Qualifications..." was offered at the 1981 Annual Conference of the International Personnel Management Association Assessment Council (Abrams, Laguardia,

22

& Sheibley, L.). The workshop included recommended characteristics of good MQs and procedures to achieve these characteristics.

An earlier training course on this topic was developed and is periodically offered by the Mid-Atlantic Personnel Assessment Consortium (MAPAC). MAPAC is an organization of East-coast states, cities, counties, and federal agencies which work on cooperative projects to improve their personnel assessment methods. One of the most successful areas of MAPAC activity has been the development and conduct of training programs to upgrade the skills of Personnel Assessment Specialists. The MAPAC training course which covers procedures for establishing job-related minimum qualification standards was developed in 1980.

The MAPAC training materials on Minimum Qualifications (MQs) includes: a review of court cases, reasons for minimum qualifications, legal standards for establishing MQs, characteristics of "good" MQs, and a content validity methodology for establishing minimum qualifications. The training includes a workshop in which participants apply the methodology.

The MAPAC review of legal cases "...seems to indicate that general education requirements (high school or college degree MQs) have generally not held up well. Experience requirements have fared somewhat better. This may be because defendants seem to have a great deal more 'content validity' or job relatedness evidence to offer for specific experience requirements than for generalized education and experience requirements" (Sheibley, L., 1980). The MAPAC procedure for determining appropriate MQs includes a worksheet for linking each entry-level important knowledge, skill and ability requirement to specific optional ways in which applicants can demonstrate possession of each requirement. The information from this worksheet is used as the basis for developing the MQs.

Development of Job-related Minimum Qualifications

Brief examples of public-sector jurisdictions which have established programs and procedures to assure job-related MQs follow.

Delaware. The state of Delaware is using a unique approach to minimum qualification requirements. Traditional training and experience requirements are generally being replaced by knowledge, skill and ability requirements. Minimum amounts of education or experience are not being established for many state positions. Instead, for each knowledge, skill or ability requirement (KSA), applicants are asked to supply information which demonstrates that they possess the requirement. A variety of types of evidence is accepted from applicants to show that they meet MQ standards.

There is a successful precedent for this procedure. In the 1950s, Ernie Primoff of the U.S. Civil Service Commission successfully developed KSA-based minimum requirements for blue collar jobs in the federal service. The innovation of the Delaware project is the attempt to apply this method to all occupational areas.

Delaware is dropping many traditional education and experience requirements because of such problems as:

- "Minimum qualifications often reflect desirable qualifications or approximate the qualifications of the incumbents providing the data rather than specifying the minimum requirements necessary…;"

- "Agencies would advocate exaggerated qualifications . . . in order to maximize the pay grade."

- "Discrimination—degree and licensure requirements tend to eliminate minorities who traditionally have had less access to these programs. Years of experience tends to eliminate women who are more likely to be in and out of the labor force" (Delaware State Personnel Office, 1988, p. 2).

Delaware has separated the process of pay setting from establishing minimum qualifications to address the problem of using MQs to justify more pay. A job analysis based procedure has been established to develop KSA-based MQs. This procedure is described in Delaware's "Knowledge, Skill and Ability Based Minimum Qualifications: A Guide for Developers."

In some cases, Delaware does not evaluate applicant qualifications against MQs. In these instances, Delaware uses a test as the sole applicant screening procedure.

Written criteria have been prepared by Delaware to limit the circumstances where degrees, diplomas and licenses are included in minimum qualifications. For example, degrees, diplomas and licenses can be required when:

— they are an absolute legal or professional requisite,

— they are the only practical means by which a required knowledge is obtained, and

— there is empirical evidence that individuals who do not possess the requirement cannot successfully perform the job.

The information used to establish KSA based MQs is documented for each job class.

The major advantages of Delaware's system are: applicants who may possess the KSAs needed to do the job are less likely to be screened out by the MQ evaluation process than with more traditional MQs; the MQs are established using a job analysis based procedure and thus should be more job-related; pay considerations should not inflate MQs; and the resulting MQs should be less likely to have adverse impact.

The disadvantages of Delaware's procedure are: the cost, since more developmental effort is required; the increased difficulty of reliably determining who meets the MQs since the MQs will likely be longer and more complicated; and the greater burden which will be placed on the later parts of the selection process since a larger group of applicants with more varied backgrounds will need to be assessed in the next part of the selection process.

To address the potential lack of reliability, Delaware is developing rating guides with a variety of examples for use by application reviewers.

Delaware is nearing completion of their work in converting minimum requirements to this new system. We look forward to an evaluation of the Delaware alternative to traditional education and experience requirements.

Pennsylvania. The state of Pennsylvania has, through management directive, issued standards for the "Develop-

ment and Validation of Classification Standards." Training has been provided to those who establish MQs in how to meet the standards. In essence the standards are designed "To ensure that information in classification standards is job-related and that this job-relatedness is demonstrable" (Commonwealth of Pennsylvania, 1982). The standards cover such topics as: job analysis, sampling, data to be collected, describing work, determining relative importance of KSAs, deriving MQs from KSAs, verifying the information, documenting the information, and the required review and approval process.

A guide has been prepared to assist in the review of draft class specifications. The guide describes a review process which helps to obtain valid and job-related minimum qualifications standards (Johnston, 1982).

Once MQs are established, applications are reviewed to determine which candidates meet the MQs. The Pennsylvania State Civil Service Commission has prepared a detailed guide for the evaluation of application information and the determination of applicant eligibility. The guide establishes policy in such areas as: defining acceptable experience and training, determining equivalencies, evaluating experience, evaluating part-time experience, crediting volunteer experience and correspondence training, accepting General Educational Development (GED) and College Level Examination Program (CLEP) scores in lieu of diplomas and college credits, degree equivalency policies, and other topics (Kehew, 1982).

Summary and Recommendations

The public sector has taken the initiative and provided training on and a technology for the development of content valid job requirements. Qualifications developed following these procedures are more likely to be fair to job applicants than traditional procedures for developing minimum qualification requirements. Guides and procedures have also been developed for reviewing and evaluating applications and resumes in a manner which considers alternative ways for candidates to demonstrate possession of minimum requirements. Examples of such efforts have been described in this article.

The National Commission on Testing and Public Policy should recommend that employers use content validity based procedures in determining their minimum qualification requirements. Such procedures can increase the appropriateness and fairness of job requirements, and reduce adverse impact. This recommendation can be followed without large expenditures by employers.

RATINGS OF EXPERIENCE AND TRAINING (E&Ts)

Background Information

In the public sector, the assessment method known as the Rating of Experience and Training (E&T) is one of the most frequently used testing procedures. A survey reported "...E&T ratings were used more than any other selection device except written tests" (Cook, 1980). Other authors note that "Although it is generally believed that the interview is the most frequently used selection procedure, T&E evaluations are actually used more often, because a person's training and experience is usually reviewed at the interview (Levine & Florey, 1975), and is often used as a screening device prior to the interview" (McDaniel & Schmidt, 1985, p. 1). A 1980 survey of major state, county, and municipal merit systems found that the E&T procedure is used much more than written tests. For example, in Tennessee of 2,000 examinations only about 10% require paper and pencil tests. "When we have relatively fewer applicants to be examined and positions to be filled we tend to use relatively simple T&E rating schemes..." (Johnson, Guffy & Perry, 1980).

In recent years, the exploration for alternative selection procedures under the provisions of the 1978 Uniform Guidelines on Employee Selection Procedures may have increased the use of the E&T method even beyond the previous heavy reliance on it.

Quantifying a person's background by rating experience, training, accomplishments, responses to questions and other data provided on application forms and questionnaires is much more common in the public sector than in the private sector. This is because of the laws and regulations surrounding hiring in civil service systems. Since government hiring procedures are open to public scrutiny, scores are usually assigned to candidates to help determine who will be considered for job openings.

Research on E&Ts

Prior to the 1980s, there was little research evidence on the validity of the E&T methods and the research evidence which did exist was not very positive. "Studies we were able

to locate concluded almost unanimously that statistically significant validity coefficients are rare, and most frequently they are zero" (Johnson, Guffey & Perry, 1980).

Considerable research in the 1980s has examined the reliability and validity of E&T methods, as well as provided data to evaluate some of the basic assumptions which underpin this method. Much of this research has been conducted in the public sector. In addition, a variety of guides and training programs have been prepared and conducted to improve the use of E&Ts. The guides, such as those provided in seminars by the International Personnel Management Association Assessment Council, the Mid-Atlantic Personnel Assessment Council, and the U.S. Civil Service Commission's guide on the behavioral consistency method, provide procedures for conducting job analysis, developing E&T procedures and rating scales, using subject matter experts to ensure job-relatedness, and other steps to establish content validity of the E&T procedure.

As an indicator of the recency of work on E&T methods, the most effective E&T method, according to current research, is the Behavioral Consistency Method. This method was first described in detail in an "Exam Preparation Manual" as "...a new examining method representing a significant improvement over the Commission's traditional methods of rating applicants in unassembled examining" (U.S. Civil Service Commission, 1977).

Prior to 1980, practically all research data on E&Ts was in unpublished government reports. There was little criterion-related validity research evidence to support the use of E&Ts prior to the 1980s, since the most frequently used method up to that point in time, the point or traditional method, had little evidence of criterion-related validity (Caplan & Schmidt, 1977).

The major research reviews on E&Ts and the various E&T methods are all very recent and are contained in the following reports: "The Validity and Utility of Alternative Predictors of Job Performance" (Hunter & Hunter, 1983), "A Meta-analysis of the Validity of Training and Experience Ratings in Personnel Selection" (McDaniel & Schmidt, 1985), and an article on "Validity Research Evidence for the

T&E Evaluation Methods" in the *Participant Manual* of the
International Personnel Management Association's Seminar
on "Ratings of Training and Experience" (Abrams, Ash, &
Johnson, 1986). All of these documents rely heavily on
public sector data.

"A Literature Review and Meta-analysis of Education as
a Predictor of Job Performance" (Dye & Reck, 1988)
provides additional relevant data.

Types of E&Ts

Prior to summarizing some of the key findings of these
research reviews, we will briefly describe some rating of ex-
perience and training examination methods. Following is a
list of most of the methods. For more details on and a com-
parison of methods, interested persons can attend an IP-
MAAC Seminar on E&T ratings and obtain the *E&T Semi-
nar Participant Manual*. The following listing is too brief to
fully describe or compare the methods. It is intended to pro-
vide summary information to those who are unfamiliar with
this assessment method, so that the information presented
later in this report is meaningful.

- Traditional Point Method. Points are assigned to ap-
  plicants based on such factors as the number of
  years of experience and training, and the relevance
  of the training and experience to the job being
  sought.

- Improved Point Method. A job analysis is conducted
  to identify important job requirements and specific
  things that applicants could have done in the past
  which would indicate possession of those require-
  ments. Specific job tasks or assignments performed,
  community achievements, courses completed,
  grades, etc. may be credited. Scoring procedures are
  developed to credit this applicant information. Sup-
  plemental questionnaires may be developed to col-
  lect the information from applicants.

- Grouping Methods. Applicants are sorted or divided
  into a small number of groups based upon an overall

30

simultaneous evaluation of their training and experience.

- Behavioral Consistency Method. A job analysis is conducted to identify important job requirements (usually from five to seven). Supplemental application forms are designed to collect information from applicants on their achievements on each important requirement being assessed. Applicants complete supplemental application questionnaires describing their achievements. Anchored rating scales are developed by job experts using actual examples of achievements for each requirement being assessed. The completed applicant questionnaires are rated by comparing responses to the rating scales.

- Self-rating E&Ts. A variety of methods fall under this category. All start with the conduct of a job analysis to identify tasks or requirements to assess. Structured questionnaires are developed to collect information on applicant qualifications for each task or requirement. For example: applicants may be asked to describe or check off tasks they have performed and indicate their degree of proficiency in performing each task or their degree of independence in performing the task; or applicants may be asked to rate the extent to which they possess each job requirement (knowledge, skill or ability).

Validity Research Findings

Hunter & Hunter (1983) compared the average validities of a wide variety of measurement methods. The methods they reviewed in the E&T category generally had low to moderate mean validity compared to the other assessment procedures reviewed. For example, "experience" had a mean validity of .18 (based on 425 studies), "academic achievement" .21 (based on 17 studies), and "training and experience ratings" .13 (based on 65 studies) for predicting performance in entry-level jobs. For predictors in higher level jobs, they found a mean validity of .49 (based on 10 studies) for the behavioral consistency method.

McDaniel & Schmidt (1985) conducted a meta-analysis of validity data on E&Ts. The study was based upon data from 124 coefficients and 11,503 cases. They found that the validity of E&Ts varies considerably based on the E&T method. They report low mean validity for the point method (.15 based on 81 studies); moderate validity for different types of self-rating methods (.23 for an Illinois job element method based on 16 studies, and .32 for a task method based on 9 studies); and highest E&T method mean validity for the behavioral consistency method (.49 based on 10 studies). All of the validity data by method, except for the point method, was found to be generalizable. Since the number of studies on many of the E&T methods was limited, the authors caution that additional research studies are needed.

An update of the above 1985 study is reported in *Personnel Psychology* (McDaniel, Schmidt & Hunter, 1988). The update is based upon 132 coefficients and 12,048 subjects. Some additional questions were examined in this more recent meta-analysis. The mean validity for the point method was low (.11 in 91 studies); the mean validity for the task method was also low (.15 in 10 studies); moderate for the Illinois job element method (.20 in 16 studies); and high for the behavioral consistency method (.45 in 15 studies). Additional reviews in this article indicated that the point and task methods "...may show useful levels of validity for samples with low mean experience levels." Based on the data, the authors concluded "...the Illinois job element and behavioral consistency methods appear to display validity generalizability" (p. 305).

A 1988 U.S. Office of Personnel Management review of the effectiveness of one type of data sometimes used in E&Ts, college grade point average (GPA), found a meta-analysis correlation of .29 between GPA and training success, and a correlation of .21 with job performance. For educational level, the study found an average correlation of .28 with training performance and .11 with job performance. A conclusion from this study was: "...on the average, grades exhibit a positive, moderate relationship with performance in the work place. Level of education is less predictive, but nonetheless is related to successful job performance"

(Harris, 1988). These findings are generally more positive than that found in previous reviews of the validity of GPA for predicting job performance.

The OPM study also examined use of information on the quality of educational institutions in personnel selection. Available evidence suggests that institutional quality measures do not have merit for use in personnel selection. A review of the literature found low validity for institutional quality measures, and no evidence that institutional quality moderates the validity of college grades (Dye & Reck, 1988, p. 15). The study, entitled *A Literature Review and Meta-analysis of Education as a Predictor of Job Performance,* provides valuable information on the relationship of a variety of educational measures to both job performance and measures of training success.

Another meta-analysis of the relationship between GPA and job success found mixed results (Bretz, 1989). Thirty-nine published validation studies were analyzed using the Schmidt-Hunter meta-analysis technique for summarizing results across studies. A new study of 328 recent graduates from bachelor's and master's degree business programs was also conducted. The author concluded from these analyses:

"The overall analysis suggests that no relationship exists while subgroup analysis of success in business and in teaching suggest that significant relationships do exist.... Based on the relative weakness of these relationships and confidence in the overall meta-analytic results, it is suggested that if a relationship does exist between GPA and job success it is tenuous at best. Since other more significant predictors of success are available, the use of GPA in this capacity is not recommended" (Bretz, 1989, p. 11).

Dye & Reck suggest that the conclusions reached by Bretz on GPA as a predictor of job success need to be reconsidered. They question Bretz's use of the effect size statistic in his analysis, the combining of the results of studies which used different measures of job success, and the statistical test used to interpret the study results. They summarize the findings of their meta-analysis which found validity generalization evidence for GPA in predicting job and train-

ing success. (Dye & Reck, 1989, pp. 235-241). Dye & Reck do not challenge the Bretz statement that other more effective predictors are available.

The findings of the research studies summarized above are having the effect of moving many public employers away from traditional point E&T methods towards self-rating methods, behavioral consistency methods, and, in the federal service, increased use of GPA as an assessment method. In addition, research to assess these methods as well as develop alternatives, is continuing.

Comments on E&T Methods

Job applicants tend to accept the E&T examination method. In general, public employers receive relatively few applicant complaints about this method, despite the low validity of the point method, which is used heavily.

E&T methods tend to be most appropriate for jobs requiring prior preparation, and less appropriate for entry-level jobs.

It is likely that the point method, with its heavy reliance on years of education and years of experience, would have adverse impact on females and minorities who are only recently entering many occupational areas in reasonable numbers. The International Personnel Management Association Assessment Council's *Instructor Manual* for the Rating of Training and Experience seminar contains these comments on the traditional point method:

> "...the traditional point method should not be used. Its apparent level of measurement is not appropriate, it results in unwarranted adverse impact, it requires nearly double the scoring time of grouping and task-based methods (Ash & Levine, 1985), its use is extremely difficult to justify on the basis of content validity alone, and empirical validity for the method appears to be near zero" (Abrams, Ash, Johnson, 1986).

There is some information, as noted later in this report, that self-rating methods and the behavioral consistency method do not have the adverse impact problem noted for the traditional point method.

34

One reviewer of this paper expressed concern that inflation of ratings may be a problem for self-rating methods (R. Reilly, correspondence, May 2, 1989). Unfortunately, the meta-analysis data on E&T validity mixes different types of self-report instruments. One type of instrument asks applicants to report on fairly objective data such as if they have performed a particular task, what equipment they have operated, what courses they have completed, or their college grades. Another type of procedure, often called self-assessment, asks the applicant to provide a rating of proficiency.

Inflation may not be a problem with well developed self-report procedures. For example, a recent study examined the relative validity of objective data on college grades and self-reported grades and concluded: "...validity is nearly identical whether information on grades is obtained through formal records or is obtained from self-reports. Previous meta-analysis have even found validity for self-reports to be twice that for objective records" (Dye & Reck, 1988, p. 40).

The behavioral consistency method depends considerably on the writing ability of the applicant and should not be used unless writing is an important job requirement.

The writing required to complete a supplemental application can discourage candidates from applying because of the considerable time and effort required to provide the information on achievements. Thus the quantity of applicants may be affected by use of the behavioral consistency method. However, one study of applicants for three professional job classes found no general or consistent relationship between the quality of applicants (as measured by subject matter expert opinions of applicant qualifications) and applicant behavioral consistency questionnaire return rates. The study also found no differences by sex or race in questionnaire return rates (Lawton, 1982).

All of the E&T methods which have shown moderate to high levels of validity take considerable time and expertise to develop and carry out. Note that these E&T methods are in effect tailor-made tests for the specific job class for which an employer is assessing candidates. As such, the developmental effort required can be similar to that needed for

many other tailored assessment procedures, such as structured oral examinations or work simulations. However, the administrative costs may be lower depending on the E&T scoring procedure.

Examples of E&T Use

Use of Grade Point Average (GPA). The U.S. Office of Personnel Management (OPM) has recently (1988) proposed a policy of allowing use of GPA as one candidate assessment method for on-the-spot hiring for entry-level college positions. In this instance, grade point average is proposed to be used on a pass/fail basis. Candidates with GPAs of 3.0 and above are to be hired without tests. This use of GPA is related to the OPM finding, reported earlier, of a meta-analysis correlation of .21 between college grades and job performance. OPM also studied group differences in grade point averages of minorities and whites. The study was based on a sample of 4,500 federal incumbents in professional, administrative and career occupations. The results of the study indicate that the proposed policy of permitting the hiring of candidates with GPAs of 3.0 or above would result in less adverse impact than the use of ability tests (U.S. OPM, J. Gandy, personal communication, 6/15/89).

Candidates will also have the option of competing through another testing procedure which includes a written test and a bio-data form called an "Individual Achievement Record" which is described in the section of this paper on biodata.

There appears to be a number of considerations which influenced OPM's plan to use GPA as an assessment method. Part of the OPM rationale is probably to address the finding of a 1988 report to the President and the Congress, "Attracting Quality Graduates To The Federal Government" which identified a need to "...shorten and simplify the competitive recruitment process while preserving the underlying merit principle" (U.S. Merit System Protection Board, 1988). The OPM program is also designed to "...enable us to obtain a quality work force which is also reflective of the population it serves" (U.S. OPM, June 23, 1988, p. 1). The Work Force 2000 report also helps to explain the OPM policy. The

projected shortages of candidates for entry-level professional positions has led OPM to attempt to simplify and streamline the assessment and appointment process. Litigation over the previous PACE alternative was probably another related influence on the policy decision.

There is controversy about this planned use of GPA. For example, the president of the National Treasury Employees Union complained to the House Civil Service Subcommittee that "...Direct, on-the-spot hiring is not open, fair or competitive" (Linsley, 1988). The 1989 meta-analysis conclusions by Bretz reported earlier also raises questions. There are many assessment procedures available with higher levels of validity than GPA; however, all are more expensive and more time consuming to administer than use of GPA. In addition, the applicant shortages OPM is concerned about may not be a problem for all of the job titles where GPA use is proposed. OPM's research also indicates that use of GPA in the college major may have a higher level of criterion-related validity, and could have a higher level of content validity than use of four year college GPA.

The use of GPA by OPM is an example of an assessment method being used to address a variety of goals such as organizational goals of merit selection and utility, social goals of a representative work force, and the goal of a quick, flexible recruitment and selection process. In this instance OPM appears willing to trade-off a likely loss in utility by using a less valid selection method for a likely gain in work force representation and a streamlined recruitment and selection process. It would be interesting to review the OPM utility analysis which provides the basis for this planned use of GPA.

The 1989 recommendations of The National Commission on the Public Service appear to support OPM's planned use of GPA. Commission recommendations included simplifying the hiring process, increasing representation of minorities in the federal service, and providing greater flexibility to federal hiring agencies (National Commission on the Public Service, 1989, pp. 5, 20, 29, & 30). The planned use of GPA could help to accomplish these recommendations.

Self-rating E&Ts. Although self-rating methods are listed here as one category of E&Ts, the reader should keep in mind that there are a variety of self-rating methods. In addition, self-rating E&Ts differ from self-assessments (e.g., self-assessment of level of job performance) although some meta-analysis studies do not differentiate between these procedures.

The self-rating method appears to have been originated by Ernest Primoff of the U.S. Civil Service Commission. He developed this method as part of the job element examination procedure, and first used it for blue collar trades occupations.

The Seattle Regional Office of OPM reported on a computerized E&T using a self-rating procedure. The procedure was developed because of cutbacks in the assessment staff while applicant volumes remained high. The computerized self-rating procedure resulted in cost reductions from 76% to 98% over the previous manual procedure.

Applicants self-rate their skill level on a set of job tasks. A variety of steps were taken to help assure accurate self-ratings, such as: requiring applicants to substantiate their claims, informing them that their former teachers or supervisors may be contacted to verify their skill ratings, and building checks into the computer scoring. There may be from 200 to 300 tasks rated for an occupation. Two of the occupations for which this procedure was used are Electronics Technician and Nursing Assistant (Long, 1984).

"Thomas J. Lyons conducted several studies at the U.S. Office of Personnel Management on self-rating T&Es. In a study of a self-rating computer T&E for clerical jobs with a sample of 6,846 applicants, he found 18% or 1,242 rated ineligible; the 5,604 rated eligible had a mean rating of 85.68 with a standard deviation of 5.44. Of these applicants a follow-up predictive validity study of 421 showed a validity coefficient of .21 as compared to .249 for the written test. These results indicate that, despite expected problems, both validity and variability among applicants were produced with this method. Lyons produced similar score distributions with groups

of applicants for a variety of Accountant and Auditor positions" (Abrams, Ash, Johnson, 1986).

A study of Public Health Nurses by the State of South Carolina to develop a self-assessment instrument was described in an IPMAAC conference session on "Training and Experience Evaluations and Other Forms of Self-Assessment" (Lacomba, 1983). A task-based structured job analysis questionnaire was developed and administered to a stratified sample of 630 nurses. Data collected on each task included: importance, entry-level, and frequency of performance. Factor analysis of the ratings was conducted to identify job dimensions. The reliability of the questionnaire was estimated as .96, and eight job dimensions were identified which accounted for 74% of the variance in ratings. The profile information of tasks associated with each job dimension will allow candidates to report their experience in a manner which "…would enable supervisors to quickly identify a candidate whose experience most closely resembles the work that is performed by the incumbents in a particular unit" (Lacomba, 1983, p. 86).

Behavioral Consistency Use. Karen Olivia White provided a summary of federal agency use of the behavioral consistency method as a result of the U.S. Office of Personnel Management delegation of examining authority to many federal agencies under the Civil Service Reform Act of 1978. One example was the Federal Deposit Insurance Corporation (FDIC) which reported receiving 400 to 500 applications each time examinations for Bank Examiner were announced. About 117 applicants were hired. After use of the behavioral consistency method, FDIC reported an unusually low percentage of ineligible applicants which was attributed to: 1) the "realistic job preview" which this examining procedure provides, and 2) applicants may pay closer attention to the job requirements before going through the process of completing the required forms. Rating time was about 15 to 20 minutes per applicant. Use of this examining method along with delegation of examining to the agency resulted in a higher percentage selection of females, Blacks, and Hispanics as compared to PACE, the previously used written test (White, 1982, pp. 18-19).

The state of Wisconsin has developed and uses an Achievement history questionnaire which is based upon a methodology similar to behavioral consistency procedures.

A New E&T Method

The "Activity Achievement Indicator" is a new E&T method developed by Ash. The method was designed to obtain scores similar to the behavioral consistency method without the problem of applicants having to complete long narrative questionnaires. "Substantial portions of applicant populations which complete traditional application forms refuse to complete the behavioral consistency application supplements" (Ash, 1986, p. 326).

The method requires significantly less writing skill of the job applicant as compared to the behavioral consistency method.

The Activity Achievement Indicator (AAI) consists of sets of three statements. Applicants are asked to choose which of each of the three statements is most typical of them. The statements are similar to those used in the rating scales for scoring behavioral consistency E&Ts, and are developed using the same procedures as used in developing behavioral anchored rating scales. Each set of statements contains one "high," one "medium," and one "low" activity or achievement for a job requirement. In the developmental study of this method, for the job of "Planner," the AAI consisted of twenty-six sets of three statements.

The procedure was tried out on a sample of college students. The correlation between AAI scores and behavioral consistency scores was .58 (N = 33). The internal consistency (coefficient alpha) of the AAI was .56 (N = 86), and the interrater reliability for the behavioral consistency method was .74 (N = 33) (Ash, 1986, pp. 337-338).

Ash concluded from the study that the AAI "...may have potential as an alternative to the unpopular (with applicants) behavioral consistency approach. There is substantial room for improvement in the AAI as detailed in this study. The internal consistency reliability estimate needs to

be raised. This can be accomplished by developing additional items for each dimension, although it will undoubtedly increase the developmental cost..." (Ash, 1986, pp. 341-342).

Summary and Recommendations

The rating of experience and training is used heavily in the public sector as an assessment method. Recent research has revealed that E&T methods vary considerably in their validity and are also likely to vary in their level of adverse impact. Traditional private sector reliance on application and resume information to make selection decisions does not have the level of validity shown by the innovative methods developed and researched in the public sector.

The National Commission on Testing and Public Policy should recommend increased use of E&T methods which have been shown to have higher levels of validity and lower levels of adverse impact. The behavioral consistency method should be strongly recommended for jobs where it is appropriate. Use of typical application and resume information, such as is usually done in industry and in the traditional merit system point methods of rating experience and training, should be discouraged.

BIODATA

Background Information

This article will review what biodata instruments are, summarize public sector reviews of biodata as an alternative selection procedure, and provide examples of use of biodata questionnaires in public personnel assessment.

The use of biodata instruments in the public sector has been very limited. This is primarily because of the problems outlined in the second section of this article. However, as will be described in the last section of the article, recent innovative efforts have addressed public sector concerns with this assessment method and increased use is now occurring.

The most unique aspect of biodata procedures is that they are typically developed following a strictly empirical methodology. Job-relatedness is not normally used as the basis for selecting questions for a biodata inventory; rather, items are selected based upon the relationship between answers to biodata questions and some criterion measure, such as job tenure.

Biodata or biographical information questionnaires consist of structured multiple-choice items. The items typically request facts about the person completing the questionnaire, such as: demographic information, education, work experience, interests and social activities, habits, hobbies, family history, attitudes, values, achievements, and personal characteristics.

These inventories can be administered to job applicants in a group setting and can be machine scored. The procedure is therefore inexpensive to administer. However, biodata instruments are expensive to develop.

The research procedure used to develop biodata questionnaires typically includes selecting a large number of potential questions and constructing a trial inventory. The trial inventory is administered to job incumbents. To obtain stable data, from 400 to 1,000 or more incumbents are needed. An analysis is conducted to determine the relationship of answers on each biodata question with a job success measure, such as supervisory ratings, turnover, or sales volume. Items which relate to the job success measure are

kept in the inventory. The scoring key for the new inventory is developed based on the relationship of the answer choices to the job success measure. In addition, a cross-validation study is conducted to make sure that the validity level of the selected items remains consistent on another sample of workers. Periodic follow-up research is needed to make sure the scoring procedure and validity level do not deteriorate.

Research on Biodata as an Alternative Method

The U.S. Civil Service Commission, the U.S. Office of Personnel Management, and the New York State Department of Civil Service have all conducted reviews and evaluations of biodata as an assessment method. Much of the research data reviewed is from the private sector which makes considerable use of biodata. The results of the reviews (McKillip & Clark, 1974; Van Rijn, 1980; Hunter & Hunter, 1983; Corcione, 1985) report positive findings of validity studies on biodata instruments for a wide variety of job titles, but also report many difficulties for using biodata in merit system examining. A summary of the findings of these reviews follows. The summary covers these topics: validity, adverse impact, advantages, problems, and conclusions.

Validity Data. McKillip & Clark (1974, p. 4-6) summarize research covering a dozen different occupational areas. The jobs included sales positions, scientists, office workers, middle managers, physicians, production employees, clerical workers, military officers, and others. The studies showed personal data "...proved to be valid predictors of success...." in these jobs. They summarize, for example, work by Asher on studies where cross-validation was conducted and the biodata questions were "hard" or verifiable. Asher found "...that, of 31 validity coefficients reported in these studies, 35 percent were .60 or higher; 55 percent were .50 or higher" (McKillip & Clark, p. 2).

Paul van Rijn's review of background questionnaires (BQs), reported:

"Time after time, BQs have proven to be one of the best single predictors of future behavior (Asher, 1972; Asher & Sciarrino, 1974; Guion, 1965; Henry, 1966; Owens,

43

1976; and Schuh, 1967). England (1971) lists numerous studies to demonstrate the validity of biodata items as predictors of success for a wide variety of jobs and for a broad spectrum of criteria. These studies typically involve criteria of job tenure, salary increase, or success in training. Tenure is by far the most usual criterion, and BQs seem to have been particularly effective in predicting this criterion for clerical workers (e.g., Robinson, 1972)" (van Rijn, 1980, pp. 8-9).

A report on "The Validity and Utility of Alternative Predictors of Job Performance" includes a meta-analysis of 44 validity studies reported by Reilly and Chao in 1982. The analysis found an average validity of .44 for the biographical inventory (Hunter & Hunter, 1983, p. 12). This was the highest validity level of the eight predictors included in the review.

Hunter and Hunter also conducted a new meta-analysis of the validity of some alternatives to ability tests, and found relatively high average validities for biodata. They report average validities of biodata as follows: .37 for supervisory ratings (12 studies, 4,429 subjects); .26 for promotion criteria (17 studies, 9,024 subjects); .30 for training success (11 studies, 6,139 subjects); and .26 for tenure (23 studies, 10,800 subjects). Compared to the other alternatives analyzed, biodata usually was the second highest alternative in terms of level of validity (Hunter & Hunter, 1983, p. 18).

The New York State Department of Civil Service review of previous validity data reached these positive conclusions: "Biodata was clearly seen as the most valid predictor of performance when compared with that of other predictors such as tests of intelligence and aptitude. The superiority of biodata was found to hold for criteria of trainability and for job proficiency" (Corcione, 1985, p. 10).

The public sector reviews reported above, which found positive validity evidence supporting biodata instruments, were confirmed by a very recent review of research. The more recent review found a higher level of validity for biodata than for cognitive ability tests (Reilly & Warech, 1988, p. 80).

Adverse Impact. A presentation to the National Civil Service League on "An Alternative to Discrimination in Testing..." presented data on the validity of biodata. A biographical inventory was used in a study of 13,000 North Carolina High School students. The questionnaire predicted college academic performance with cross-validation coefficients in the mid-sixties. This level of validity was higher than that for any of 24 scores from conventional tests. The biodata instrument did not discriminate between Blacks and Whites (Ellison, 1971, p. 1).

Concerning adverse impact, McKillip & Clark caution that: "Since many biographical items deal with social and economic variables, members of disadvantaged groups might be disproportionately excluded from consideration for selection on the basis of unfair or non-job related facts" (1974, p. 7). A similar caution is stated by van Rijn: "Biographical questionnaires may have some adverse impact, most likely as a function of the extent to which the questionnaires elicit information that either directly or indirectly reflects our country's cultural differences in opportunities for economic, educational, and social advancement. Biographical questionnaires that are constructed to avoid such questions may help to reduce adverse impact" (1980, summary, p. 2).

A review of research on the adverse impact of biodata was reported by the New York State Department of Civil Service:

> "Owens, Sparks, Cherry, Cascio, and Schoenfeldt have all found little or no adverse impact in the use of biodata.

> "Jennifer French, of the San Bernadino County Employment Division, is currently experimenting with the use of biodata for entry-level clerical positions and eligibility workers. She has found thus far that biodata is the only test that has no adverse impact" (Corcione, 1985, p. 9).

The more recent Reilly & Warech review of alternatives found biodata to be one of four predictors which showed lower adverse impact than cognitive ability tests (1988, p. 80).

U.S. Office of Personnel Management research on a newly developed biodata instrument, reported on later in

this paper, also found considerable reduction in group differences on the biodata instrument, as compared to ability tests.

Advantages. The major advantage of biodata instruments, in addition to the high level of validity and low or no adverse impact, is the low administrative cost. These instruments can be quickly administered to a large group. They can be machine scored. Another advantage is that the security of biodata items is not a concern. Only the security of the scoring key needs to be maintained. An additional advantage is that biodata, because of high validity, has a lower false rejection rate than many alternatives. A smaller proportion of rejected applicants would be successful if hired as compared to most other assessment methods (Reilly & Warech, 1988, p. 84).

Problems. Biodata instruments present a number of problems. Some of these are of special significance in the public sector because of merit regulations governing assessment and selection. These problems are identified and discussed in the public sector reviews listed earlier.

One problem is job-relatedness. Merit system regulations require that assessment devices used in the public service be job-related. Usually this is assured by developing assessment devices from job analysis data. Biodata instruments have not been typically developed in this fashion. The typical biodata instrument would have problems of applicant acceptability in the public sector since the questions are usually not clearly job-related. Another problem related to this is how does one go about explaining results or scores to applicants? In the public sector, applicants have the right to be given detailed information on their test performance. Applicants want to know what they can do to improve their scores. This is difficult to do with biodata instruments since scoring key information is confidential. Release of scoring key information would probably destroy the validity of the biodata instrument.

Another problem is invasion of privacy. Many items typically asked in biodata instruments are of a personal nature. In the public sector, such questions cannot be used since they cannot be shown to be job related, they would vio-

late invasion of privacy regulations, and they would not be acceptable to candidates or public officials.

Some reviewers have cautioned that biodata questionnaires may be situationally specific. McKillip & Clark (1974, p. 8), for example, caution that changes in labor markets, populations, etc. may change the relevance of the biodata information. Reilly notes "that a recent report (Hogan & Stokes, 1989) demonstrated that biodata scoring keys developed under research conditions do not generalize at all to applicant populations" (Richard R. Reilly, correspondence, May 2, 1989). However, the Hogan & Stokes research may not provide adequate evidence on this matter. The study was based upon turnover of salesmen, used "soft" items, used a relatively small total sample in the predictive study (N = 555, including both short and long tenure), and found rather modest cross-validation levels for the biodata instrument (.23 and .22).

Fakability may be a problem. Applicants may answer biodata questions differently than the employee group in the validation study. Intelligent applicants may be able to determine which answers will make them "look good." Paul van Rijn cites a few examples of studies which illustrate this problem. He states: "...there remain enough concerns about the fakability of BQs to warrant additional research in this area" (1980, p. 10). Reilly also states "There is considerable evidence that applicants will fake or distort their responses to biodata questions when jobs are at stake" (Richard R. Reilly, correspondence, May 2, 1989).

Biodata validity may diminish over time. Hunter and Hunter cite studies which illustrate this (1983, p. 19). To address this problem, users of biodata instruments need to conduct periodic follow-up studies to make sure validity does not decay. These studies increase the cost of this assessment procedure.

There are feasibility and developmental cost problems with biodata. To effectively develop such instruments requires large samples in the range of 400 to 1,000 or more for the tryout study, and more large samples to conduct follow-up studies (Hunter & Hunter, 1983, p. 20). Many employers do not have enough employees in given occupations to con-

duct such studies. In addition, research and statistical expertise, and large scale data analysis computing capabilities are needed. The time and cost to develop and maintain a biodata instrument are considerable.

Conclusions. The reviews of biodata generally indicate that biodata has been shown to be one of the best assessment methods for a variety of jobs and criteria, with low or no adverse impact. However most public sector reviewers generally advise caution because of the many problems outlined above. Considerable ongoing research would be required to make use of biodata in the public sector. Special efforts would be needed to address unique public sector constraints in the use of biodata, such as the need to make all biodata questions job-related.

Paul van Rijn sees the greatest value of biodata as post-employment placement tools rather than selection tools (1980, p. 15). McKillip & Clark see biodata as "...a long-term, high risk effort" (1974, p. 10). Corcione expresses a more optimistic view: "The opportunities for improving selection based on biodata appears to be almost unlimited in those situations where the applicant flow is sufficiently large to permit developing an adequate data base" (1985, p. 11).

Examples of Biodata Use

Federal Government. McKillip & Clark reported on studies of biodata conducted at NASA:

> "The results showed that creativity and other scientific performance measures could be predicted by biographical data for incumbent scientists and engineers. The validities tended to lessen (although still statistically significant) when the biodata were gathered on newly hired research professionals followed up one year after appointment" (1974, pp. 5-6).

A biographical questionnaire was developed to reduce turnover in the Air National Guard. The study attempted to develop items which would not "perpetuate past discrimination" as may be done by typical biodata life history items. "Present life" items assessing current or very recent behavior were developed. The results indicated that the

"present life" items possessed greater validity than other life history items used in the biodata questionnaire. The author concluded: "The major contribution of this study is that it offers an alternative type of biodata item which should be legally defensible, have less adverse impact and have at least equal validity" (Kleiman, Falls, & Wilcox, 1982, p. 59).

The most recent innovation in the use of biodata is that of the U. S. Office of Personnel Management, which reports:

"The federal government is planning to use a biodata measure for the first time as part of their new examining program for entry-level administrative and professional jobs. The new instrument is called the Individual Achievement Record (IRA), and is an empirical measure of personal characteristics which predict successful job performance. Unlike versions which are in use in the private sector, questions for the Federal biodata measure were carefully screened to be sure they related only to events which were under the control of the individual. (This completely eliminated socio-economic variables from use.) Questions were also vigorously reviewed to avoid bias and invasion of privacy and to insure that only job-related items were included. The empirical keying was established using a sample of more than 6,000 recent hires into entry-level Federal jobs nationwide. The empirical research indicates strong validity evidence ($r = .31$, cross-validated), and significantly reduced adverse impact (effect size of .27 SD for Blacks and .09 SD for Hispanics). These encouraging results, for a biodata measure specifically designed to be appropriate for use in the public sector, are a significant step forward for sound and balanced personnel selection" (U.S. Office of Personnel Management correspondence, March 3, 1989).

A comparison of the above effort with the problems noted earlier for biodata, indicates that steps were taken to overcome many of the problems which biodata instruments have had for the public sector. OPM plans future research to follow up on those hired. A "Work Force Quality Assessment Research" program is in place. This effort should tell

if the problem of fakability and the problem of decay in validity occur for this new instrument.

The reduction in group differences with the IRA, as compared to typical paper and pencil ability or knowledge measures is also encouraging. The group differences on written ability tests are typically three to four times greater (U.S. OPM, June 23, 1988, p. 2).

Governmental Consortia. Richardson, Bellows, Henry & Co., Inc. (RBH) is conducting a consortium validation research study of selection procedures for law enforcement jobs. Participating jurisdictions include: the Suffolk County, New York Police Department; the New Jersey State Police; Warren, Michigan Police Department; the Las Vegas Metropolitan Police Department; and the Georgia State Patrol. The focus of the study is on patrol positions. One of the selection procedures in the study is a biodata instrument developed by RBH called the Law Enforcement Candidate Record (LECR).

An RBH review of previous entry-level testing programs for law enforcement jobs found that they:

"...were all uni-dimensional, specifically general aptitude-cognitive in nature. They also all had the typical substantial impact of such tests on minority and/or female subgroups" (RBH, 1988, pp. 2-3).

The purpose of the research effort is:

"...to develop a demonstrably job-related selection procedure for use in the selection of law enforcement officers, no matter the organization's geography, size, or type. It was also undertaken to determine if the use of autobiographical questionnaire scores, most likely combined with scores on other instrument types, could achieve that objective with at least equal validity but with lower score differences than those observed for the typical types of tests now in use.... It is intended in the long term to bring together the largest sample ever assembled for the purpose of law enforcement officer selection instrument development research" (RBH, 1988, pp. 2-3).

A job analysis was conducted of positions in the consortia jurisdictions. A sample of 2,088 participated.

"...the job analysis results indicate a substantial agreement across samples on law enforcement officer task requirements.... there is almost total agreement on the abilities required to perform tasks successfully.... it was concluded that sufficient empirical support existed to warrant (a) combining the individual organization samples into one total research sample, and (b), assuming a successful outcome of the investigation, the use of a common selection procedure across all organizations" (RBH, 1988, p. 13).

Six tests were used in the research: verbal comprehension, arithmetic reasoning, spatial analysis, number scanning/checking, recall/memory, and a biodata instrument.

The job success measure used as criteria in the study consisted of performance ratings collected for research purposes. Ratings were obtained from first and second line supervisors on 1) task performance, and on 2) the ability of the officers. An analysis of the rating data found "less variance than observed elsewhere" for such ratings, and low to moderate inter-rater agreement (range of .33 to .55). Another reliability study is underway to check on the stability of the ratings over time.

The biodata validity estimates (within the sample on which the biodata scoring key was developed) ranged from .24 to .35. These correlations were higher than for the other tests in the research. An analysis of the fairness of the test battery was conducted. This study was based on a sample of 16,987 candidates for police positions in Suffolk County. The Black/White mean score difference on the test battery, which included three ability tests and the biodata instrument, was 6.98 points. This difference is equal to .83 of the standard deviation for the total sample. Compared to the previous test used by Suffolk County, the number of minorities in the upper score ranges increased by 50% (RBH, 1988, p. 33). The report did not include separate data on group differences for the ability tests and the biodata instrument.

This consortia research effort is continuing.

State Government. The state of New York conducted
an extensive study on the use of biodata to improve selection
procedures for Mental Hygiene Therapy Aides. The state
employs over 21,000 employees in over 40 mental health and
mental retardation facilities in this job title. The title has a
history of high turnover and poor employee performance
(Corcione, 1985, p. 1). A written test was used to select
employees.

New York State hired Occidental Consulting Group
(OXICON) of McGraw Hill, to work with state staff on the
development of a biodata questionnaire. A two year study
was conducted. The reports reviewed on the New York
studies did not indicate if any special efforts were taken to
address the problems with public sector use of biodata which
were cited earlier in this paper. "During the first year of the
research study, a biographical questionnaire was developed
which appears to predict, without adverse impact on
protected class members, the high performance and long
term tenure probability of candidates..." (Corcione and
Means, 1986, p. 50).

Corcione's description of the first year study illustrates
many of the administrative problems associated with con-
ducting a large concurrent validation study. Unions as well
as supervisors resisted the study and made it difficult to con-
duct (1985, pp. 104-105).

A second-year predictive follow-up study reported cor-
relations of .14 with performance evaluations, and .20 with
turnover. The correlation of the performance biodata key
with a combined performance and turnover criteria was .18.
The correlation of the employment status biodata key with a
combined performance and turnover criteria was .19. Al-
though the correlations were statistically significant, "...Con-
siderable doubt was expressed by Civil Service
Testing staff as to whether these correlations were of any
practical significance" (New York State Department of Civil
Service, 1988, p. 1). One problem noted in the study was
low reliability of the performance criterion.

Based on the follow-up study results, the New York
State Department of Civil Service decided not to operational-
ly make use of the biodata instrument.

Local Government. Corcione's review reported on this use of biodata for clerical positions in county government:

"Jennifer French, San Bernadino County Personnel Division, has been using biodata as a weighted portion of their written examination for entry-level clerical workers. Ms. French indicated that their use of biodata for clerical workers was an unqualified success, and was the single best predictor of any part of the examination. Ms. French found biodata had the highest, most significant correlation with performance. She also added that biodata had no adverse impact" (Corcione, 1985, p. 3).

Use of biodata for entry-level county firefighters was reported by Carol Bullock. The counties were Anne Arundel, Baltimore, and Howard County, Maryland. The report did not include data on the results of the research (Bullock, 1983, p. 76).

Research on the development and use of a content-valid biodata instrument to predict turnover for Municipal Police Officers was reported by Walter Mann. He found a statistically significant correlation of .33 between the biodata score and turnover over a two and a half year period. Mann stated, "It is possible therefore that content-valid BQs do not have to be revalidated as frequently as BQs developed using a criterion-related approach. Research is needed to study the deterioration in the predictive validity of content-valid BQs over time" (Mann, 1983, p. 78).

Summary and Recommendations

The above examples indicate that there is hope for the use of biodata in the public sector, especially when special efforts are taken to address the unique problems which this technique presents for the public sector. The New York State experience indicates that we need to be cautious in our expectations. The federal government appears to have committed itself to the needed long term research to determine if the problems can be overcome. The generally found high level of validity and low adverse impact of this procedure warrant this research commitment.

The National Commission on Testing and Public Policy should recommend continued research on biodata instruments, given the high levels of validity and low levels of adverse impact of well researched and carefully developed instruments. Consortia activities to pool resources in order to conduct needed research should be encouraged.

STRUCTURED ORAL EXAMINATIONS

Background Information

This article provides summary information on the use of structured interviews in merit system examining, summarizes research findings on use of the interview generally, reports on recent findings demonstrating the superior reliability and validity of structured oral examinations, and provides recent examples of use of structured interviews in the public sector.

Structured oral examinations are most appropriate for use in assessing job requirements which can not be readily measured with paper and pencil tests such as: oral communication abilities, interpersonal abilities, and supervisory and managerial abilities. As noted by reviewers of the first draft of this report, use of a structured oral examination to only assess job knowledge would be an expensive procedure, especially with a large applicant group. With small applicant groups, however, use of a structured oral examination can be less expensive than development of a job knowledge test and can also assess many other important job requirements.

The employment interview may be the most common selection procedure in use, along with the application form. In the public sector, a 1974 survey by Huett and Ford found 93% of responding organizations using panel or oral board interviews (structured oral examinations). A private sector survey on the use of the interview by 852 firms in 1957 found 99% using it (Huett, 1976). Davey (1984) commented on the high acceptance of the interview by applicants and managers.

One major difference between public and private sector use of the interview is the heavy use of structured interviews in merit system examining in federal, state, and local government. For example, in 1975 the Pennsylvania State Civil Service Commission gave structured oral examinations to 3,457 candidates for 95 job classes (Moreano & Sproule, 1976). This study demonstrated unusually high levels of reliability for the structured civil service oral examination. Of 133 inter-rater reliability coefficients on 51 different oral

examination boards, the median was .90 and 80% were above .80.

A 1988 Personnel Program Inventory Survey of International Personnel Management Association agency members found that 76% of the responding 389 public sector jurisdictions and agencies were using structured oral examinations. The structured oral examination was the selection procedure used by the largest percentage of respondents. Twelve types of selection procedures were included in the survey questionnaire.

Huett's 1976 report on "Improving the Selection Interview in a Civil Service Setting" outlines an oral examination procedure which has been standard practice in many public sector organizations. The procedure includes defined steps for the development and conduct of oral examinations, such as:

Conduct a systematic job analysis to identify the important job elements.

— Develop a standard set of questions linked to each of the important job elements.

— Use questions which elicit factual, verifiable information on actual behaviors.

— Develop evaluation forms for use by interviewers in recording and summarizing their observations.

— Develop precise standards of evaluation.

— Communicate the important elements of the job to the interviewers.

— Use multiple interviewers. Train interviewers in questioning techniques, use of rating forms, and how to avoid errors.

— Have interviewers rate candidates independently. Avoid comparisons of candidates to one another.

— Evaluate the results of interviews and take action to correct inconsistency.

A recent International Personnel Management Association Assessment Council Monograph also contains many research based recommendations on steps to improve the selection interview (Silverman, S. & Wexley, K., 1987).

Previous Research on the Interview

Previous research on the interview generally found low levels of validity. This is probably because most interviews studied were individual unstructured interviews, which may not have been developed using procedures to assure job relatedness.

Huett (1976) reported on previous research on the interview. He summarizes this research as generally finding low inter-interviewer reliability and low validity. His report includes a listing of 51 validity studies from as far back as 1916 which included 53,649 subjects. The average (mean) uncorrected validity was about .20 (Davis, 1986).

Davey reports:

"...the validity of the typical employment interview is dismal. Possibly the most thorough inquiry into the validity of the employment interview was that undertaken by Ulrich and Trumbo (1965), indicating generally low validity for this method. Tenopyr (1981) reports finding only one published validity study on this topic between 1978 and 1980, while finding approximately fifty articles warning of possible defects and pitfalls published during this period" (1985, pp. 366-367).

A review of validity data published by the U.S. Office of Personnel Management (Hunter & Hunter, 1982) reported a meta- analysis mean validity of .14 for the interview for entry-level jobs for which training will take place after hiring. This average was based on 10 correlations and 5,389 subjects. The report includes a meta-analysis of 5 correlations of the interview with promotion as the criterion, the average validity was .08. For training success, an average validity of .10 is reported for the interview based upon 9 correlations with 3,544 subjects.

Huett (1976) identified these problems with previous interview procedures: failure to use standard interview practices, failure to use standard procedures for summarizing observations, use of ambiguous rating scales, failure to provide job data and information on applicants on which to base conclusions. The structured procedure outlined earlier summarizes some of the steps Huett recommended to overcome these deficiencies.

In evaluating the meaning of previous interview research, many reviewers have reached the same conclusion as Huett. One review (Arvey & Campion, 1982) expressed hope for the structured interview. Davey (1984) also expressed the view that the structured oral examination is a special case which "...promises a higher level of validity than the traditional interview by virtue of its use of multiple raters, standardized questions and structured scoring."

Perhaps the main reason for the low validity reported in past research on the interview is that most interview studies were of unstructured, unstandardized individual interview procedures, which were not based upon job analysis or clearly linked to job performance requirements.

New Research on Structured Oral Examinations

Recent research findings support the hopes expressed above by Huett, Davey, and Arvey & Campion. The reliability and validity of structured interviews is considerably higher than that previously shown in the literature for the interview generally. It is interesting to note that much of the research on the structured oral examination is conducted in the public sector.

A report by the U.S. Office of Personnel Management (Whetzel, McDaniel & Schmidt, 1985) analyzed interview reliability and validity data by type of interview using meta-analysis procedures. Job related interviews were better predictors than psychological interviews. Structured interviews were better predictors than unstructured interviews. The highest levels of validity were found for structured interviews. The highest expected validity generalization result was .51 for job related structured interviews with job performance criteria collected for research purposes. This was based on 10 correlations with 978 subjects. Using typical "on file" performance rating criteria, the validity generalization result was an expected correlation of .26 based on 46 correlations with 6,376 subjects. Using training success criteria, the validity generalization result was .24 based on 7 studies with 1,161 subjects.

As stated by Whetzel, McDaniel & Schmidt, p. 36, "The conclusions drawn in this paper are in sharp contrast to the traditional beliefs held about the validity of the interview." The likely reason for this contrast, as stated earlier, is that previous interview research was primarily on individual unstructured interviews which may not have been job related.

A report on "Raising the Validity of the Oral Examination" by the Chief Examiner of the King County Civil Service Commission (Davis, 1986) lists ten situational interview validation studies with an average validity coefficient of .28 (uncorrected). These include four studies by Davis and six studies by Latham. The highest levels of validity were on three managerial studies, where the mean validity was .38 (uncorrected).

Davis believes that validity is highly content-based and to improve the validity of the interview one needs to add more content to it, or improve the content such as by making the content more relevant to the objective (e.g., job mastery). He states that situational interviews derive their content from job analysis, usually via the critical incident technique. Situational interviews pose problems to candidates about job performance related situations. He believes the distinguishing characteristic of situational interviews is in use of the BOSS technique: Behavioral Observation Scales. These are the evaluation criteria by which candidate responses are evaluated. They are, in effect, the test answer key. He recommends designing such tests in terms of the desired answers.

A worldwide literature search on the validity of the employment interview located 150 studies which met six criteria established by the authors. A meta-analysis of the 150 studies was conducted. The results were that "...structured interviews produced mean validity coefficients twice as high as unstructured interviews." For the 48 studies (10,080 subjects) which were based on structured interviews, the mean validity was .34 uncorrected and .62 corrected (Wiesner & Cronshaw, 1988, p. 284). The authors state: "...The validity coefficients of structured interviews both individual and board, are comparable with the best

predictors available to industrial/organizational psychologists, including mental ability tests" (p. 286). "The results also suggest that higher validity coefficients are associated with more reliable interviews and the use of formal job-analytic information in developing interview questions" (p. 275). The results for the 9 studies (1,328 subjects) where the interview was based on a formal job analysis was a mean validity coefficient of .48 uncorrected and .87 corrected (p. 287). An additional interesting finding of this study was that consensus board ratings were more predictively valid than statistical combination of independent ratings (p. 288).

A recent meta-analysis of data on the structured interview (Wright, Lichtenfels & Pursell, 1988) found an estimated mean validity of .39 based upon 13 studies with 827 subjects. The uncorrected validity coefficient was .27 on 14 studies with a total sample of 870. This research combined data from six new validity studies of the situational interview with seven previous studies by Latham. The authors note (pp. 15-16) that the validity of the structured situational interview approaches that found for many cognitive ability tests.

Wright, Lichtenfels & Pursell (1988, pp. 4-5) state that the superiority of the structured interview over the unstructured stems from three factors: 1) structured interviews are based on a thorough job analysis, 2) structured interviewing assumes that individual intentions and behaviors are strongly linked, and 3) structured interviews use the same questions for each applicant with predetermined standards for answers.

The 1988 Annual Review of Psychology includes a chapter on Personnel Selection and Placement (Guion & Gibson, 1988, p. 367) which reports on a 1987 large sample study by Arvey of many interviewers in many locations. "Uncorrected validity coefficients of .34 and .51 were reported for the use of a structured, scored form in two different years."

A recent International Personnel Management Association Assessment Council *Personnel Assessment Monograph* on "The Selection Interview: Some Reasons for Optimism" also reports recent positive findings on the use of

job-related, structured interviews. (Silverman & Wexley, 1987, pp. 4-5).

The U.S. Office of Personnel Management recently reported on the results of a series of validation studies to identify alternative selection procedures. The research compared the relative validity and adverse impact of six different selection procedures by a meta-analysis of recent federal validation studies.

The results were summarized in a paper presented at a symposium, "The Federal Government's Search for Alternatives to Traditional Ability Tests." The findings follow.

"The structured interview performed very well and seems to offer the best opportunity for reducing group differences." The report notes that use of the structured interview alone, without use of "a good cognitive ability test," would likely result in a loss of utility. "These results support a strategy of test development which seeks to optimize combinations of tests, one of which should be a general cognitive ability test" (MacLane & O'Leary, 1988, pp. 7-8). The authors note the costs involved in using structured interviews might be considerable.

A state-of-the-art research study on structured interviewing was recently reported in *Personnel Psychology* (Campion, Pursell & Brown, 1988). This study compared the reliability, predictive validity, utility, and fairness of a structured interview to paper and pencil tests, using a job performance measure developed for research purposes as the criterion. The sample was 149 entry-level production employees hired from 243 interviewed. A battery of four written tests was administered to those hired before they started work.

Some of the results of the above study included a corrected validity coefficient for the structured interview of .56; an estimated utility of over $1 million in ten years from use of the structured interview; and no test fairness problems. The reliability and validity of the structured interview was comparable to the expected mean validity of cognitive aptitude tests. (Cognitive aptitude tests have had the highest expected validities in meta-analysis studies comparing measurement methods). The authors outlined the reasons why

use of structured interviews would warrant the additional expense involved, compared to use of written paper and pencil tests. The article also describes recommended procedures for the development and conduct of structured interviews.

Some of the research reported above, as well as some of the research in the next section of this article (which gives examples of recent public sector interview use), contain data which indicate that the adverse impact of structured oral examinations is considerably less than that for paper and pencil assessment tools. This may be due to such factors as: the care taken in the selection and training of interviewers, especially efforts to avoid bias; the loading of the interview on listening and speaking abilities as opposed to reading and writing abilities; and the variety of responses which candidates can give to demonstrate their qualifications.

In summary, recent research indicates that well-developed, carefully administered structured oral examinations, based on job analysis information, using job-related questions, specific and anchored rating scales, and well-trained raters have high reliability and a level of validity comparable to that of cognitive ability tests, and show less adverse impact. The criterion related validity coefficient in recent studies are much higher (a range of .24 to .62 in the studies reviewed above) than previous research findings on interview validity. The structured oral examining technique has been standard practice in public sector merit systems for decades.

Examples of Structured Oral Examinations

Following is a sampling of reports on public sector use of structured oral examination. The reports show the variety of occupations where this assessment tool is used, provide further data on reliability, validity and adverse impact, and illustrate procedures used.

One innovation which occurs frequently in public sector use of the structured interview is to provide candidates with case problems or study materials prior to the oral examination. Some jurisdictions provide a time period (e.g., one hour) prior to the oral for candidates to study problems and

questions. This practice is closer to on-the-job situations than the straight question and answer interview procedure. This form of the structured interview has some similarity to work simulation testing.

Candidate acceptance of the structured interview is high and challenges are relatively few. However, the cost, time and expertise needed to properly develop and carry out structured oral examinations is considerable.

U.S. Office of Personnel Management. "The Office of Personnel Management has utilized the structured interview extensively as components of examinations for selected occupations for several years. Examples of their uses range from entry level to more senior positions, such as Professional and Administrative Career entry level, Presidential Management Intern Program, and the Administrative Law Judge selection process. The interview provides an excellent assessment of knowledges, skills, and abilities which are oriented towards 'meet and deal' activities involving interpersonal skills. Each interview includes several job-related questions which simulate critical incident situations encountered on the job. To facilitate scoring of the applicant's responses, a set of benchmarks and standards has been developed by subject matter specialists regarding appropriate responses. Each interviewer is required to receive a minimum of three hours of training, using a specialized tailored course book, accompanied by videotapes of a simulated interview, for practice scoring. In some programs, the structured interview has been considered as an oral test portion of the total examination. It has proven to be a necessary component for certain types of examinations and is valid for certain purposes and typically has very little adverse impact" (U.S. Office of Personnel Management, Office of Personnel Research and Development, correspondence, 3/3/89).

U.S. Department of Labor–Employment Service Interviewer and Managers. Nettles (1981, pp. 42-43) reported on "Developing and Designing Reliable and Standardized Interviews." The report was on a U.S. Department of Labor funded project for the Employment Service titles of Interviewer and Local Office Manager. The selection proce-

dure used for both jobs included written objective tests and standardized interviews designed to assess tasks or knowledges, skills and abilities. Research was conducted using video taped interviews to determine reliability. Nettles reported these results:

"In all, 54 panels of interviewers observed and scored the three simulated interviews. Interclass reliability coefficients were computed for each panel. The lowest coefficient obtained was .69. Ninety percent (90%) of the reliability coefficients exceeded .85, and 80% exceeded .90. The ratings of the three interviews were remarkably similar over all groups.... These results seem to confirm that, when steps are taken to standardize the content of interviews and the method of evaluating or rating the results is carefully scaled, the scores or results of interviews can be quite reliable."

Interpreters. Ferrara (1981) described the merit selection procedure used in the New York court system for the hiring and placement of Spanish-speaking Court Interpreters. The process includes a written test and a structured oral examination. Procedures for the oral included the development of standardized examination scripts derived from actual court transcripts and task-based job analysis information, the development of behaviorally-anchored rating scales, examiner training sessions, and other steps to insure standardization. One hundred forty candidates were examined in the oral. Reliability results obtained showed that "...raters can make accurate decisions when there are specific criteria and an opportunity to observe actual behavior that is expected in the target position."

Fire and Police Promotional Tests. Use of the critical incident technique for development of oral examinations in Fire and Police promotional tests in Arlington County, Virginia was reported on by Janet McGuire (1982 p. 45). The procedure included high involvement in the process by managers from within the departments. Objective scoring check lists were developed. "Fire and Police reaction to the procedure was strongly positive. The technique had good interrater reliability and tended to correlate more highly with management ratings of performance than did the

64

written examination which had also been used in the promotional process."

Police Promotion Examination. Flynn, Anderson, and Rubovits (1984) conducted a reliability study of a structured oral examination procedure used as part of a police promotion examination in a small northeastern city. The examination was developed from job analysis data. Fifty-two candidates participated in the oral examination. They were independently rated by four raters. An alpha reliability coefficient of .91 was obtained.

Police Selection and Promotion. Davey (1984) reports on the use of structured interviews in police employment selection in the State of Connecticut. He found a predictive validity coefficient of .23 across six oral examination panels which examined over 700 police candidates, 121 of these were hired. He found some differences in average scores by rating panel despite the high structure of the examining process.

Davey and Wallace (1986) reported on a study of a structured oral examination for Police Sergeant. The paper reports on an attempt to address the problem of differences across oral examination rating panels by building a high degree of structure. The oral was administered to 352 candidates by three oral boards. The oral was developed based on job analysis data. Standard questions and rating scales were used. For eight of ten questions in the oral, answer keys were established which consisted of elements which should appear in candidate responses. Points were deducted if responses did not include the elements. The two other questions were scored using a Likert-type scale. Guidelines were provided for scoring these questions. Candidates were provided a 300 page study guide two weeks prior to the examination.

Extremely high inter-rater reliability resulted from the above procedure. The average reliability coefficient was .95. This correlation was based on final board member ratings after discussion of independent ratings; however, there was a very low (10%) rate of change after discussions. Analysis of the ratings revealed that the high rate of structure also seemed to reduce halo error in the ratings. Another finding

was that there were differences in spread of scores and mean scores between the boards. This was addressed by standardizing the scores of each board.

Corrections Officer Selection. Berkley (1984) studied the predictive validity of a multi-part examination program for selecting entry-level Corrections Officers in the Pennsylvania Department of Corrections. One part of the examination was a structured oral conducted by a three member board using procedures similar to those described earlier by Huett. An uncorrected predictive validity coefficient of .29 was found for the structured oral examination for 180 persons hired. This is the correlation between total oral test score and job performance at the end of the first 12 months on the job as assessed by "on file" supervisory ratings. The written test predictor correlated .16 (uncorrected coefficient, significant at the .05 level) with the same criterion. The written test combined with the oral test in a multiple regression equation did not predict job performance any better than the oral examination alone. Final candidate grades, corrected for restriction of range and criterion reliability, correlated .43 with the performance ratings.

The Correction Officer selection research by Berkley (1984, p. 46-51) also examined many other relationships and questions, such as the utility or value of the selection procedure. He found that with use of the test, for each 100 hires, nearly 10 dismissals would be avoided. Over 200 hires occur per year. Berkley used equations from Schmidt et al. to estimate the dollar value of the selection procedure. He found that "...use of the examination, rather than random selection, to hire Corrections Officer Trainees would result in a productivity savings (over the long run) of over $7,000,000 for each year the test is used." It is interesting to note that the major contributor to validity in this research was the structured oral examination.

Adverse Impact

Abrams (1985 pp. IV-33) reported on the adverse impact of oral examinations.

"The literature on the adverse impact of all types of oral examinations is confusing. All processes allow for rater bias to influence the results. A number of studies indi-

bias to influence the results. A number of studies indicated that female and minority candidates received lower ratings, especially from white male raters. More recent studies have indicated that, with heightened sensitivity for affirmative action, raters may rate women and minority candidates higher than white males (Teske, 1984) and (Iverson and Teske, 1982)."

The study on Police Sergeant noted above (Davey and Wallace 1986) found no adverse impact for a structured oral examination.

Obtaining female and minority oral examination board members, and training board members on how to avoid bias has become common practice in public sector structured oral examining.

Many public sector jurisdictions have prepared guides or training materials to help insure fairness and proper development and use of interviews. One example is the State of Washington booklet on "Improving the Interview" (Valadez, 1987), and "Oral Examination Development Manual" (Schultz et al., 1983). The state of Washington, like many public jurisdictions, also offers a training course for supervisors and managers on how to prepare for the interview and how to conduct it.

Summary and Recommendations

Considerable use of structured oral examinations takes place in the public sector. Actions to refine and improve this assessment method are common in the public sector. The procedure is used for a variety of jobs and job requirements. Structured oral examinations can assess many important job requirements which cannot readily be assessed by paper and pencil tests. Use of the procedure, however, can involve considerable administrative costs for large applicant groups. Public sector assessment professionals have prepared guides, training materials and procedures which have increased the validity of interviews.

Recent research indicates that well-developed, carefully administered structured oral examinations, based on job analysis information, using job-related questions, specific and anchored rating scales, and well-trained raters have high reliability and a level of validity comparable to that of cognitive ability tests, and show less adverse impact.

The National Commission on Testing and Public Policy should recommend the structured oral examination methods developed and researched in the public sector to all employers. Such procedures are a vast improvement over typical employment interviews and are likely to help reduce adverse impact.

Since employers are likely to continue using interviews as part of their employment procedures, the cost of converting current unstructured interviews to job-related structured interviews is not excessive, given the likely consequence of high validity and low adverse impact.

WORK SIMULATION, WORK SAMPLE AND PERFORMANCE TESTS

Background Information

In recent years, there has been considerable activity in public sector development and use of work sample, work simulation, and performance tests for personnel selection. For example, one assessment specialist reported, "The trend in this jurisdiction for some time had been a movement towards the use of work sample approaches to testing..." (McGuire, 1986).

Such approaches include a wide variety of assessment methods which sample what employees are required to do on the job. The rationale for these assessment methods is essentially: "If you want to know how well anyone can do anything, the closer you get your test to match what it is you are trying to predict, the more accurate your prediction will be" (Joiner, Dennis A., 1987, p. 30).

These assessment procedures are understandable to the test taker. When well developed, their job relevance is usually obvious. The recent evidence of validity of such tests is high. Courts have looked favorably on work sample tests. In addition, such tests typically have less group differences than paper and pencil ability tests.

The major disadvantages of work sample assessment procedures are the developmental effort required, and the administrative time and costs associated with conducting and scoring them. In general, work sample tests tend to be more job specific than most published paper and pencil tests. This increases the costs of such tests since their use is more limited than some other tests.

Users of work sample tests must consider "...whether the more complicated technique is yielding any information over and above what could be gained by less complex techniques" (Reilly, Richard R., personal correspondence, 5/2/89). Data and information related to this concern can be found throughout this article.

One recent trend to control costs has been the use of a small number of simulations to assess the most impor-

tant parts of the job. As reported by one assessment specialist: "A recent trend in the federal government has been to make greater use of the assessment center methodology, but to use fewer and less complex exercises in order to be more cost effective" (Baker, 1988).

Ingenuity and creativity on the part of test developers can help to simplify the process and overcome the disadvantages, as illustrated in many of the examples from the public sector in this article. Some jurisdictions have used creative approaches to incorporate work simulation exercises into structured oral examinations, paper and pencil tests, computer administered tests, take-home tests, and other examination formats. Machine scoreable in-basket tests, latent image testing, and a "take-home" test are other examples of innovative work simulation tests described in this article.

There is a high level of activity in the development and use of work sample tests in the public sector. This indicates that many public sector jurisdictions believe that the advantages of these testing methods outweigh the additional developmental, administrative and scoring costs such tests can require.

Types of Work Sample Tests

Some examples of work sample, work simulation, and performance tests follow.

Psychomotor skill tests, such as:
- Typing
- Drafting
- Vehicle operation
- Machine operation (sewing machine, word processing, duplication machine)
- Tool use
- Trade skill (welding, painting, engine repair, electronics repair, plumbing, carpentry)
- Air traffic control simulation
- Airplane flight simulator operation

- Cardiopulmonary resuscitation

Physical Performance Tests, such as:
- Ladder climbing, hose carrying, hose coupling (Firefighter)

Supervisory and Management Ability Measures, such as:
- In-basket exercises
- Leaderless group discussions
- Interview Simulations (e.g., employee counseling)
- Management/Business games
- Analysis problems (e.g., budget, organization, priorities, fact finding, decisions on alternatives)

Writing Exercises
- Written problem analysis
- Preparation of legal brief or quasi-judicial decision
- News release
- Letter to client or appellant

Interview Simulations
- Client interview to obtain or provide data

Oral Communications Exercises
- Report to representatives of the public or the press
- Sales presentation
- Radio vehicle dispatch exercise
- Conduct a staff meeting

Command Simulations for Fire and Police positions, such as:
- Riot control
- Emergency fire situation
- Hazardous chemical spill

Trainability Exercises
- Replication of a training task (e.g., how to process a claim form, or give mouth-to-mouth resuscitation)

71

The list of examples of work simulation tests illustrates that this assessment procedure can be used for a wide variety of job requirements, job levels, and job types. In addition, a wide variety of techniques, technology, equipment, materials, personnel, and so forth can be used in the development and administration of such tests.

Validity Data

Plumblee described 1970s research results on the "Role of Performance Tests In Improving Personnel Selection":

"Experience with the use of performance tests has varied. Cronbach (1970, pp. 388-394) cited evidence that performance tests may add little beyond paper and pencil tests for complex jobs which require intellectual skills. However, Cascio and Phillips (1979) provide evidence of cost effectiveness, lack of adverse impact and substantially reduced turnover for a wide variety of performance tests for selection in a city government. A study by Schmidt, et al. (1977) showed that a performance test for machinists had less adverse impact than a written test. Asher and Sciarrino (1974) reviewed available validity studies of work sample tests and found that motor work sample tests (involving the manipulation of things and equipment) were more often predictive of job proficiency than were other types of tests studied and were surpassed only by biographical data. Verbal work sample tests (primarily language or people oriented) were lower than motor in predicting proficiency but rated better than several other standard predictors. The level of validity coefficients reported suggests that the work sample test would be considered either as an alternative or in its own right" (Plumblee, 1980, pp. 2-3).

An analysis of "The Validity and Utility of Alternative Predictors of Job Performance" found work sample tests to have the highest mean validities for predicting job performance in promotional situations. The mean validity of work sample tests was .54. In discussing the various types of. tests compared, the authors stated: "...these predictors are

essentially predicting future performance from present or past performance. By a tiny margin...the work sample is best." For entry level jobs, the authors found ability tests to have the highest mean validity (.53); and "job tryout," which is the ultimate performance test, had the second highest mean validity (.44) of eleven predictors which were compared (Hunter & Hunter, 1983, p. 22).

A draft report on "The Validity and Fairness of Alternatives to Cognitive Tests" found work sample tests, assessment centers and trainability tests (which are special types of work sample tests), to be three of four identified testing alternatives which "...have demonstrated equal or greater validity than cognitive ability tests with less adverse impact in some situations" (Reilly & Warech, 1988, p. 86).

In summary, recent reviews of validity data indicate that a high level of validity is typically found for work sample tests, and such tests typically have low or no adverse impact.

Examples of Guides on Work Simulation Tests

A number of public sector publications in the late 1970s and in the 1980s illustrate the interest and activity in work simulations.

Minnesota Manual on Assessment Simulations. The Selection Research Unit of the Minnesota Department of Personnel prepared a 120 page report on "Job Specific Assessment Simulations" with the support of an Intergovernmental Personnel Act grant administered by the U.S. Office of Personnel Management (Farrell, 1979). This report includes a manual for developing job-specific assessment simulations as well as one complete case study example of a simulation. The report also gives examples of a variety of simulation exercises and links them to the types of job requirements they can best assess. According to this report:

> "The job-specific assessment simulation is a prime example of a viable compromise between easily developed selection procedures that are inadequate measures of competency, and adequate job-related approaches that require a great expenditure of time and effort (e.g., assessment centers). Although simulations may involve

greater commitment of time and resources compared to other selection methods, the additional costs appear to be outweighed by possible benefits derived from employing a content valid device with predictive potential" (Farrell, 1979, p. 39).

U.S. Office of Personnel Management Guide. The U.S. Office of Personnel Management published "A Short Guide To The Development Of Work Sample and Performance Tests" (Plumblee, 1980). This guide describes a step-by-step process including such topics as: planning the test development project, choosing a testing technique, job analysis, planning the test, defining the test tasks and procedures, establishing rating procedures, tryout, analysis of tryout results, final test preparation, test analysis, and validation.

IPMAAC Monograph. The International Personnel Management Association Assessment Council published a *Personnel Assessment Monograph* on "Content Valid Testing For Supervisory And Management Jobs: A Practical/Common Sense Approach" (Joiner, Dennis A., 1987). This publication describes various simulation and work sample testing methods, including: working from the job analysis information to construct a job-related examination process, types of exercises and what they can assess, the advantages and disadvantages of the procedures, typical administrative procedures, time typically allowed, general guidance on use of the procedure, and other topics. A variety of exercises are described.

The monograph is rich in examples of exercises for supervisory and management jobs, as well as examples of "public safety field command exercises." The importance of simulating the job as closely as possible is stressed. "...Simulate several of the most important situational contexts of the job to obtain the best overall assessment" (Joiner, p. 28). The author of this Monograph has also produced a videotape on exercises in the fire service. The video tape illustrates a fire command simulation exercise, a press conference exercise, and a subordinate counseling exercise.

Examples of Work Simulations Use

Following are examples of public sector use of work simulations for: upward mobility jobs; clerical and administrative jobs; trades jobs; jobs requiring communications skills; jobs requiring translation skills; entry-level professional and technical jobs; supervisory and management jobs; and police, fire, and corrections jobs.

Upward Mobility Examples. A report by the U.S. Office of Personnel Management describes the development of a work sample approach to selection for Claims Technicians jobs. These are positions which provide a career opportunity for untrained persons to advance within the federal government. Persons in these jobs review federal employee benefit claims to determine if the claims should be paid.

The test was a miniature version of tasks which employees would perform on the job. All information needed to perform the tasks was contained in the test. In the miniature training and experience work sample test, applicants demonstrate their ability to learn how to perform job tasks. They apply what they have learned to actual job tasks. They do not need to have prior knowledge of how to perform the task, as in the usual work sample test.

In this study, the test was developed following the content validity model. The study describes the development and use of the test and concluded: "The work-sample approach to selection for upward mobility was supported by the results of this study" (MacLane, 1979, p. 18).

MacLane described the results of state level use of the miniature training and experience work sample test for an upward mobility occupation as follows:

> "Criterion-related validity data are available for a very carefully developed entry-level test for Public Assistance Workers in Pennsylvania. The items in this test followed the training-and-evaluation model. Some of the tasks required in this job were similar to tasks which the Claims Technician will be doing. They involve for example, doing simple arithmetic, interpreting information for consistency with legal guidelines, and reviewing information for accuracy.... Data for the Public Assistance Worker test...showed that the traditional aptitude

measure and the work-sample test were related to a training performance criterion and were not significantly different from each other. When the relationship of these measures with current performance ratings were examined, however, the work sample test was significantly related to the ratings...the aptitude test was not.... The obvious job relatedness of the work sample test was also cited as an advantage by these authors. In sum, there is logical and empirical evidence that a training model work-sample measure is a technically appropriate selection instrument which has particular advantages for use with applicants for upward mobility jobs" (MacLane, 1979, p. 3).

Clerical and Administrative Assistant Examples. In a paper on scoring work samples, McGuire described the use of work sample tests in government for three different types of jobs.

Job A was for entry-level clerical workers in a tax office.

"...People in these jobs performed detail-oriented desk work, adding and checking figures, processing forms, and handling correspondence related to taxes. For two hectic months, they also staffed crowded information windows, and responded to long lines of angry, confused taxpayers, many of whom had unusual names, spoke limited English, or were unfamiliar with the state's tax procedures.

"For Job A, the work sample format selected was a set of tasks similar to the primary duties of the position. The examination process included an alphabetizing exercise, an exercise involving standard forms and letters, a tax form checking and correction exercise, and an interactive role play exercise. Each exercise represents an assignment that every new employee would face, with little training, during the first two months of work. They were chosen because they represented assignments where poor performance would lead to an immediate consideration of terminating the employee if they could not handle the assignment" (McGuire, 1986, p. 3).

Job B was a journey level word processing job. The examination process included both a written practical test of job knowledge and a work sample exercise. The work sample included: 1) a draft letter with hand written notations which was to be entered into the word processor as a new document and printed, and 2) a three-page report on the word processor which was to be edited and printed.

Job C was a position in an elected official's office. The need was for someone who could handle a wide variety of written materials and make appropriate judgments on sensitive and complex issues in the absence of the official. A clerical in-basket exercise was developed.

"A resource folder was compiled for candidates including a simplified list of office policies and responsibilities, several schedules and routing lists, and other materials to provide guidance for the exercise. This resource folder was provided to candidates in advance and kept by them for reference during the exercise. Items in the in-basket included correspondence, mail, notes from the supervisor, items to prepare and type such as meeting agendas, replies to correspondence, and phone messages to handle" (McGuire, 1986, p. 9).

Trades Examples. The state of New Jersey Department of Personnel conducts what is perhaps the most comprehensive performance testing program for trades positions in the public sector.

"It began in 1960–with a problem. Our agency (then the Department of Civil Service) was receiving numerous complaints from labor organizations. Paper and pencil testing for trades titles was a disservice to their members who could perform effectively on the job, but because their verbal skills were limited, failed our tests" (Fales, 1989, p. 1).

Initially, New Jersey developed trades performance tests which could be administered "out of a suitcase" or from the back of a pickup truck at an exam site. After the first successes with this procedure, the first permanent home for the trades testing effort was obtained; an unused 30-horse

stable. Later, the trades testing center was moved to an abandoned motor pool garage; and, still later, a building was designed and built for the purpose of trades testing.

Some of the jobs for which New Jersey has conducted performance tests are: Bridge Painter, Maintenance Repairer, Roofer, Floor Tile Installer, Water Main Repairer, Vehicle Operator (truck, backhoe, bulldozer), Body and Fender Repair, Carpenter, Spray Painter, Welder, Sheet-metal Worker, Auto Mechanic, and Bricklayer.

According to the Supervisor of the New Jersey Trades Examination Section, the advantages of the performance testing approach includes the ability to:

> "...replicate or simulate actual job performance in work that seldom requires significant verbal fluency. Thus test development is tailor-made for content validation... It is the simplest and most economical type of test to prepare and administer for trade and trade-related positions that yields the most desirable result" (Fales, 1989).

The metropolitan government of Nashville and Davidson County, Tennessee, provide another example of the use of work simulations in their recently developed selection procedure for their Equipment Operator positions. Traditional paper and pencil tests were not identifying better performers and a new procedure was developed. The procedure includes a training program to teach candidates the principles of equipment operation, routine maintenance and general safety; and a performance testing program to evaluate candidates for equipment operation jobs. Training programs and work sample tests were developed for these pieces of equipment: rubber tire loader, dozer, backhoe, tandem truck, and refuse collection truck, Faust (1988).

Jobs Requiring Written Communications Skills. A detailed example of the development of a work simulation examination to measure various aspects of the job of Information Officers in the state of Minnesota is described in a report on "Job-specific Assessment Simulations" (Farrell, 1979). The report includes copies of the simulations as well as scoring and administration procedures. A variety of exercises were developed for three Information Officer job titles. The exercises took four hours to administer for the job title

involving the most testing time. Other assessment devices were also developed.

The work simulation exercises included:

- writing a news release, radio spot announcement, and 30-second television commercial based on information and instructions provided,
- planning and designing a publication,
- preparing a publicity campaign for a newly enacted firearms registration law,
- making recommendations for redesigning and improving a newsletter, and
- an in-basket exercise.

The report also includes an evaluation and data on these job-specific assessment simulations. For example, the interrater reliability was high (.83 to .94). The author states: "...with a moderate investment of time and resources, job-specific assessment simulations have a great potential for improving and augmenting existing selection programs" (Farrell, 1979, p. 40).

A further example of the use of work simulations in assessing worker communication skills is found in a report by the Wisconsin Department of Employee Relations on "Writing Assessment for Merit Employment." That simulation provides information on writing skills tests for four occupations: Personnel Specialist, Property Assessment Specialist, Budget and Management Analyst, and Architect/Engineer. The study illustrates the progress being made in rater training and in the development and use of rating scales and instructions to candidates which improve the reliability of writing assessment. The study also includes a comparison which shows the cost advantages of using an essay format in measuring written communications skills for occupations with small candidate populations (Lawton & Showers, 1980).

Additional examples of work simulation writing skills assessment can be found in the section of this paper on the use of video technology. One such example is for the job of Parole Hearing Officer. Another is for the job of Corrections Officer.

Jobs Requiring Translation Skills. An example of the use of a work simulation test to assess oral and written language translation skills is described in the section on "Use of Video Technology in Assessment" in this paper. The example is from the New York State Office of Court Administration for the position of Court Interpreter (Minter, 1988).

"The U.S. Office of Personnel Management has developed and used several different types of work simulation measures as selection procedures. The most widely used of these is the Miniature Training and Evaluation (MT&E) test. The MT&E is an instrument which incorporates the training of individuals to perform a sample of relevant job tasks and the measurement of their ability to perform the tasks upon immediate completion of the training. It is based on the concept that applicants who can demonstrate the ability to learn and perform samples of essential aspects of the job in the condensed training and testing situation can also learn to perform successfully on the job, given adequate full-scale training later. One form of the test which OPM has pioneered has been developed for entry-level professional and technical jobs which rely heavily on the use of cognitive skills. Applicants are provided a training booklet containing exercises which competitors must learn within a reasonable time limit. A test booklet contains a set of test questions for each section of training. The questions are usually simulations of actual job tasks or samples of work adapted to a test format and designed to measure knowledge gained from the training booklet. The MT&E has compared favorably to traditional cognitive ability tests in studies in which the two types of instruments were involved" (U.S. Office of Personnel Management, Office of Personnel Research and Development, correspondence, 3/3/89).

Supervisory and Management Jobs–Assessment Center Exercises. An article in the U.S. Office of Personnel Management's "Operations Letter" described the federal government's extensive use of Assessment Centers, which are "a form of work sample" tests.

80

"Since 1970, OPM has been actively involved in assessment center activities with governmental agencies. The assessment center method is one of the more acceptable and valid evaluation procedures used in the identification and selection of individuals for supervisory and managerial positions. It has also been used extensively in selection for upward mobility programs and management intern programs.... Research has indicated that this method is neutral in terms of race and gender. There are sufficient data available which show that assessment centers are unbiased in their predictions of future performance. In general, the courts have viewed assessment centers as valid and fair.

"OPM is currently providing on-site technical assistance to FAA...for selecting employees with supervisory potential to become Air Traffic Supervisors...OPM is providing technical assistance to the U.S. Postal Service in updating and modifying their executive assessment center." OPM is also developing "...prototype simulation exercises for several levels, from upward mobility to executive ranks. These job-related exercises can serve as components of examinations used in the identification and selection of individuals for higher level jobs..." (Baker, 1989).

A March 1989 report from the U.S. Office of Personnel Management on work simulation tests for management jobs, states:

"Another innovative type of work simulation measure developed by OPM is the mini-assessment center. It uses the same process as the full-scale assessment center but has fewer exercises. The results provide evidence of actual behavior, but on a lesser number of abilities. Examples of this more cost effective approach are the Presidential Management Intern Program and the OPM Management Development Program" (U.S. OPM, OPRD, correspondence, 3/3/89).

Two jurisdictions, the state of New Jersey and New York City, have developed an efficient procedure for the administration and scoring of in-basket tests. The in-basket

test is the assessment center exercise which typically has the highest level of criterion related validity for predicting managerial success. In addition, the in-basket assesses many of the dimensions identified as most important for managerial success.

The New Jersey Department of Civil Service has developed multiple-choice in-basket tests for five different management occupational areas, including social service, engineering, and accounting. The procedure was developed to allow testing of large candidate populations in an efficient manner.

As with traditional in-baskets, candidates receive a variety of materials to work with, such as organization charts, correspondence, reports, messages, background information, budget information, disciplinary matters, and so forth. Multiple-choice questions were developed on the actions one could take with the problems presented by the in-basket. Correct answers were arrived at by consensus of a group of management consultants from within state government.

More than 400 candidates took the first test of this nature in New Jersey. Candidate feedback was positive and appointing authorities are requesting increased use of the procedure based on results to date (Kraus, 1986).

The New York City Department of Personnel has also developed and used multiple-choice in-baskets. The procedure is viewed as a "...cost-efficient/effective instrument amenable to the standardized scoring of a large to very large candidate population" (MAPAC News, July 1988, p. 4). Dr. Catherine Cline of New York described the procedure as a "Thurston Scale-value" approach.

> "Given a stimulus (such as an item from a typical in-basket exercise such as a memo, phone message, report, etc.) a large number of possible actions was developed. A number of judges familiar with the job, environment, etc. are then called to sort and rank the possible actions according to some scale (for example, 'decisiveness'). The judge's rankings are then subjected to...standard statistical analysis, and those possible actions are selected which form a self-consistent monotonic

continuum with respect to the scale of interest. These scale points or anchors are then assigned values to reflect their 'desirability' or 'non-desirability' with respect to the scale" (MAPAC News, July 1988).

The New York City multiple-choice in-basket was used as part of a promotion examination for a job class with over 4,000 positions. The test was administered in three-and-a-half hours in a manner similar to any objective test. Candidates recorded their answers on a machine readable answer sheet. The New York City Department of Personnel is quite satisfied with the results achieved so far with this efficient work simulation testing procedure for managerial jobs.

Police, Fire and Corrections Assessment Exercises. There is considerable public sector work simulation testing activity in selection and promotional evaluation programs for police and fire positions.

In a 1984 review of "Assessment Centers in the Public Sector" Joiner reported on the use of Assessment Centers in California State government and in California municipalities. He reports:

"The implementation of assessment centers in public agencies was aided by Intergovernmental Personnel Act (IPA) grants such as the one awarded jointly in 1974 to the California State Personnel Board and West Valley College.... The purpose of this grant was essentially to determine whether it would be practical from a cost benefit basis to implement assessment centers as an ongoing part of the selection systems of government agencies in California. The project...was quite successful. By the end of the 16 month grant period, four assessment centers had been implemented in California municipalities for the assessment of potential at the target level of Police Sergeant, three programs for assessment of potential for the Police Lieutenant level, one for Assistant Police Chief and Police Inspector and one program for Supervising Traffic Inspector, a high level management position in the California Highway Patrol. Evaluation of these assessment center programs by the involved candidates, law enforcement and personnel

management professionals indicated a high level of acceptance for the process.

"Building on the success of the project, the California State Personnel Board and others continued using and adapting the basic model which developed out of the grant. The model has now been used successfully for all promotional ranks of police and fire services as well as a variety of non-public safety classifications including County Personnel Director, City Engineer, Public Works Director, Municipal Transit Manager, Chief Building Official, Parks and Recreation Director, and Affirmative Action Officer" (Joiner, 1984, p. 437).

Similar extensive use of assessment center procedures for police command positions is also common in local governments in Florida.

The New Jersey Department of Personnel has been developing an assessment procedure for Fire Lieutenant/Captain. The procedure is being designed to have less adverse impact and be more job-related than the previous examination. Development began with a detailed job analysis in 12 cities in 1986. A series of exercises have been developed for use in candidate assessment. The exercises include:

- three fire scene simulations,
- a leaderless group discussion,
- a personal interview, and
- a training exercise.

"In the Fire Scene Simulations...1486 candidates responded in essay format to three fire scenes. The first and second scenes—which required action as a pumper captain—contained four questions. The third fire scene—which required action as a truck/ladder captain — contained three questions. The fire scenes were simulated by photograph(s) and diagram(s), with instructions on equipment available, conditions, etc." (Gordon & Siracusa 1989, p. 6).

In the leaderless group discussion five or six candidates met as a Fire Department Committee to establish a

questionnaire on supervisory style. The committee had 45 minutes to complete recommendations.

In the personal interview, each candidate had to meet for 15 minutes with a problem firefighter. The purpose of the meeting was to discuss the job performance of the employee.

The training exercise required each candidate to train a new firefighter in the operation and maintenance of a Self-Contained Breathing Apparatus. Candidates were given time to prepare for the training and 15 minutes to conduct the training session.

Take Home Work Simulation. The city of Louisville has recently used "take home" examinations for two jobs: Chief of Fire Prevention and Fire Training Officer.

> "Fire Training Officers are required to research, develop, and present training programs for the Louisville Division of Fire. The 'Take Home' gave candidates two weeks to research and develop a comprehensive written training outline on Arson Investigation Techniques for Company Commanders. At the end of the two week period, candidates were required to present a 15-30 minute training seminar composed of one part of their outline. The presentation was videotaped and both the written outline and oral presentation were graded by Subject Matter Experts" (Jeff Prewitt, personal correspondence, 2/9/89).

Prewitt described these added benefits to the city from this process:

— the Fire Training Bureau got several high quality outlines and presentation ideas,

— the candidates got a realistic job preview, and

— the tapes from the examination were used to give the newly hired employees feedback on their presentation techniques.

Latent Image Examinations. Beginning in 1984, with the help of Personnel Decisions, Inc., the New York State Department of Civil Service developed and currently uses latent image examination procedures for police and correction positions. With this assessment technique, candidates

choose an answer to a question or problem and use a special pen to obtain instructions based upon their answer. This is a type of branching or decision-making test. This type of testing allows simulation of a variety of complex real-life problems which occur on the job. Such a test was recently developed for local Police Chiefs in New York. An earlier use was in a promotion examination for Corrections Lieutenant.

Summary information on latent image testing is reported by New York State Department of Civil Service as follows:

> "Situations are set up in an 'opening scene,' which also sets the test takers objectives or tasks. Candidates are then directed to a section of the problem in which they make one or more choices in response to the situation described. Their choices move them along to other sections of the problem in which they make more choices, usually changing the situation until it is resolved. The degree of success of the solution depends upon the degree of agreement with a number of subject-matter experts who evaluate the possible courses of action represented by the choices and assign scores to the choices based upon their knowledge and experience.

> "Candidates receive test booklets that are printed partly in regular ink and partly in latent (invisible) ink. Candidates make a choice, then 'develop' or expose the correspondingly-numbered area in the latent image column or response booklet. To do so, they use a felt-tipped marker that resembles a highlighter. The latent-image response usually tells the candidates 'what happened' based on their choice, and often where to go next. Because they get feedback this way, candidates can't change their answers once they've made a choice—it is often clear that a choice was ill advised. However, as in life, what's done can't be undone; corrected later, perhaps, but not undone.

> "Written simulation problems...measure facets of candidates abilities that are difficult to evaluate with multiple-choice tests: judgment, decision-making, applied

knowledge, 'process' thinking, and some aspects of interpersonal relations. Time constraints can be set to help evaluate ability to formulate solutions under pressure.

"These areas have usually been reserved to oral tests, but the written simulations can be more efficient, especially when the candidate field is large. At present (in New York, and as compared to orals) costs of simulations 'break even' at about 100 candidates. Above that number of written simulations generally are less costly than orals.... When the candidate field is in the 1,000s the cost differences are dramatic...

"Written simulations have been or are being used to test candidates for a wide variety of jobs, from personnel specialists and health-related program administrators, to correction sergeants and police chiefs, to human services professionals. Many people who have taken simulations have responded favorably, because it was easy for them to see the relationship between the test and the job" (New York State Department of Civil Service, undated).

A report summarizing the use of the latent image procedure for local Police Chiefs in New York stated:

"The simulations assess critical decision making and problem solving skills in the context of typical situations encountered on the job. In order to avoid administration errors, New York State developed a video to train monitors how to administer this new test. Candidates were also helped to prepare for the test. They were sent a package well before the examination date explaining the new test and had a practice problem to do at the beginning of the actual test. Results have been good enough to justify the investment of the extra resources required to develop, administer, and score the simulations. Candidate acceptance was high and statistical results indicated that the test was working as expected" (Kaido, 1988).

This testing procedure has also been used in a number of other public jurisdictions, including the city of Rochester, New York, and the state of California.

One of the reviewers of this report commented that it is incumbent upon users of complex techniques such as latent imaging "...to examine whether the technique is yielding any information over and above what could be gained by less complex methods" (R. Reilly, correspondence, 5/2/89). New York State would be able to supply information related to this question from their studies to interested readers.

Summary and Recommendations

In summary, there is extensive and innovative use of work sample, work simulation, and performance tests in the public sector for a wide variety of job levels, job types and job requirements.

Recent research evidence indicates that these procedures have a high level of validity, and result in low or no group differences as compared to typical paper and pencil tests. Job-relatedness and high applicant acceptance are other positive aspects of these procedures. The major disadvantages are the required developmental, administrative and scoring costs.

This article described many examples of public sector use of work sample tests, as well as a number of innovations to reduce costs, such as cooperative consortia efforts, "take home" tests, latent image testing, video taped tests, machine scoreable simulations, mini-assessment centers, and cost analysis to determine the break-even point for use of work sample tests as compared to other procedures.

The National Commission on Testing and Public Policy should encourage and promote the development and use of work sample tests. Innovations and research in this area of assessment should be encouraged.

Intergovernmental cooperative efforts should be encouraged to reduce the developmental costs for individual employers and to expand use of innovative procedures. Federal funding to support and encourage research and intergovernmental cooperation, such as the examples cited in this paper, should be restored to the levels provided in the 1970s through the U.S. Civil Service Commission (now OPM). The state and local section of OPM as well as OPM regional and central office assessment and research staff and functions should be restored to the levels of the 1970s, which formed the basis for many current innovations.

SELECTED FEDERAL ASSESSMENT INNOVATIONS

BACKGROUND INFORMATION

The United States Office of Personnel Management (OPM) has made or is in the process of implementing major innovations in federal assessment procedures. These include:

- revisions in entry-level testing procedures,
- test simplification,
- use of logic-based measurement, and
- a work force quality assessment research program.

This is the only section of this paper which is unique to the federal government. Other information of federal innovations is included throughout this report.

Staff members of OPM's Office of Personnel Research and Development (OPRD) wrote and submitted information on innovations and developments which OPRD believed would be of special interest to The National Commission on Testing and Public Policy. In addition, information was submitted by staff of the U.S. Merit System Protection Board (MSPB). Much of this section consists of direct quotes of the information submitted by OPM and MSPB.

ENTRY-LEVEL ASSESSMENT PROCEDURES

Professional and Administrative Career Hiring

In a 1987 report to the President of the United States and The Congress, MSPB reviewed a topic "...particularly important to a civil service based upon merit..." The topic was entry-level hiring for 118 Professional and Administrative Career (PAC) occupations formerly covered by the Professional and Administrative Career Examination (PACE).

In September 1982, PACE was abolished by OPM because of charges that it adversely affected minority groups. OPM developed new alternative examinations for 16 of the

occupations, and then authorized federal agencies to fill the remaining jobs formerly covered under PACE by use of an excepted service (Schedule B) appointment authority. In effect, this delegated recruitment and selection processes to federal agencies and eliminated centralized competitive examinations.

The MSPB reported on the results of this delegation as follows:

"Agencies that have used this Schedule B authority generally have found it adequate to meet their entry-level PAC staffing needs. Largely because it allows them more flexibility in targeting recruitment activities, those agencies have increased the representation of Blacks and Hispanics among entry-level hires. However, user agencies generally find the authority administratively cumbersome and, for them, more costly than centralized competitive examining...

"Since March 1979, OPM has developed alternative examinations for 16 of 127 occupations that, by their estimate, accounted for approximately 60 percent of annual entry-level hiring under PACE. However, even if this percentage is reflective of current hiring patterns, there are still a substantial number of occupations and a substantial percentage of anticipated hiring needs without a formal competitive examination vehicle....

"From the perspective of merit system integrity, it is unfortunate that relatively little progress has been made in developing alternative examinations in the almost 6 years since the PACE was abolished. MSPB recommends that OPM accelerate its examination development process to permit elimination of this Schedule B authority in a timely manner" (U.S. MSPB, 1987, pp. i-ii).

Another MSPB report to the President and the Congress summarized the findings of a survey of selected college and university placement officials and students on attitudes towards the federal government as an employer.

"The results of this study suggest that the government is not perceived as an employer of choice by many graduates of some of the country's most highly rated academic institutions. Furthermore, even among those graduates who have a positive view of the government as an employer, many are perplexed by the 'civil service hiring labyrinth' and find little active encouragement on the part of most federal agencies. This raises concerns about the future quality of the federal work force and its ability to effectively and efficiently carry out the necessary functions of government" (U.S. MSPB, 1988, p. vii).

The recommendations in the MSPB report included: developing alternative compensation strategies, reviving a government-wide recruitment program including on-campus recruitment, shortening and simplifying the competitive recruitment process while preserving the underlying merit principles, and a bipartisan effort to improve the public image of federal employment (U.S. MSPB, 1988, p. vii).

Findings and recommendations similar to the above are also contained in the Volcher Commission Report on *Rebuilding The Public Service* (National Commission on the Public Service, 1989).

Reports on OPM efforts which address these problems can be found in many sections of this paper such as the sections on biodata and ratings of experience and training. Reports on other OPM efforts follow.

OPM is about to begin a new process for filling entry-level professional jobs.

"In 1988, the Office of Personnel Management undertook the initiative to design and develop a new examining program for the 100+ occupations previously covered by the Professional and Administrative Career Examination (PACE). This examination had been terminated in 1982 under the terms of the Luevano Consent Decree.

"The design of the new examining program for PACE occupations had as its primary goals: 1. to institute an examination or set of examinations which would reflect the state of the science in psychometrics, and 2. to meet the seemingly contradictory principles of selecting a quality

work force and one that is reflective of our national diversity.

"The resulting design for a set of six examinations represents one of the major innovations undertaken in the Office of Personnel Research and Development in recent years, and the examinations themselves promise to be among the finest ever instituted in public (or private) enterprise.

"The six examinations were designed after a major job analysis was conducted on all PACE occupations. The purpose of this job analysis was to determine which occupational homogeneities would yield the optimum number of occupational groupings for examining purposes. Preliminary results reveal six occupational groupings. These results are being corroborated at present and should be final by spring 1989. On the basis of this analysis, job specific examinations are being developed for each occupational grouping. These groupings are: Health, Safety, and Environmental Occupations; Writing and Public Information Occupations; Business, Finance, and Management Occupations; Personnel, Administration, and Computer Occupations; Benefits Review, Tax, and Legal Occupations; and Law Enforcement and Investigation Occupations.

"Each examination will consist of an ability test, tailored to the content or subject matter of each occupational grouping, plus a biodata questionnaire, called the Individual Achievement Record. Scores on each examination will be computed on the basis of equal weights for the two parts of the examination. It is expected that, since both biodata questionnaires and occupationally specific ability tests have substantially reduced adverse impact, these examinations will be instrumental in ensuring the selection of a work force reflective of our national diversity. Also, the ability tests are based on the principles of logic-based measurement, and they have been carefully designed, in the task/ability linkage process, to be economical, nonredundant measures of the intended ability constructs. As such, these tests should

serve to enhance test validities. Furthermore, the examinations will be based on a more thorough assessment of job-relevant characteristics than had been available heretofore.

"In sum, the new examining program represents an innovation that should serve the federal government very well in meeting its policy goals of selecting a representative, optimally productive work force. It is particularly appropriate to note that these goals are being met in large part because OPM has made optimal use of state-of-the-science psychometric techniques and innovations" (U.S. OPM, correspondence, 3/3/89).

Apprentice Assessment

A 1987 MSPB report on entry-level hiring identified a new federal apprentice written examination as "a success story." The new examination was developed in 1983 and first used in 1984. The MSPB stated:

"OPM and the Department of the Navy cooperated in developing the examination.

"The federal government has some seven or eight thousand persons employed in its 4-year apprentice program; between one and two thousand new apprentices are hired each year. The examination that was previously used did not differentiate well among successful candidates, making selections difficult. Further, women taking the examination seldom ranked highly, and there was an unacceptably high turnover rate among new hires who entered federal service through the examination.

"Based on experience during the first full year of using the new examination, both OPM and agencies that employ apprentices are pleased with it. The first year results indicate better-quality selections, plus general improvement in the representation of women and minorities among selections... If continuing validation testing upholds early results, this examination will constitute a success story" (U.S. MSPB, 1987, pp. ii-iii).

MSPB compared federal hiring data for apprentices with national figures and found that the government was doing better in apprentice affirmative action than the national average with the new testing procedure, as well as improving the quality of apprentices (U.S. MSPB, 1987, p. 22).

In the 1987 report, MSPB stated that OPM was conducting a criterion-related validation study of the new apprentice examination using both training and job performance criteria. The results of that study were reported in September 1988 by Walter G. Mann of OPM.

The test consists of a nine-part written battery. The abilities assessed include: eye-hand coordination, measuring ability, form perception, complex arithmetic, memory, reading comprehension, numerical reasoning, table reading, and simple arithmetic computation. The criterion measures were status in the training program (e.g., being behind schedule, having been separated from the training program, etc.), and course grades.

The study was based upon a sample of 810 apprentices hired by the Navy from 1984-1986. The study found "…a very high validity for predicting success in apprentice training" (Mann, 1988, p. 14). Validity generalization analysis found that validity was consistent by job location and type of trade. The uncorrected correlation of the test and average course grades was .58. The report does not provide information on the relationship between test or training performance and job performance.

The parts of the test with the highest predictive levels in a regression analysis were the three test parts which had been developed especially for the apprentice jobs (reading comprehension, numerical reasoning, and table reading). Use of the regression results to improve the battery would result in six test parts, and a 23% increase in predictive power over the original nine-part test.

An analysis of the test passing score was conducted. The recommended 84th percentile passing score, "…will produce a very low failure rate with enough minority eligibles to provide appropriate representation" (Mann, 1988, p. 14).

The data from the above study should lead the MSPB to conclude that their 1987 prediction of "a success story" for this new apprentice examination was accurate.

FEDERAL TEST SIMPLIFICATION PROGRAM

One of the principal innovations initiated during the past year in the Federal Examining System is a two-pronged effort to simplify tests.

"One aspect of simplification involves simplifying test batteries by removing redundant subtests. Research carried out during the past decade, for example, in the area of logic-based measurement, has identified convergences between factors, and has permitted the refinement of measurement in certain areas. Accordingly, all subtests identified as redundant measurements in existing test batteries have begun to be removed. The resulting system will be a simplified, less cumbersome, and more economical testing system.

"A second aspect of simplification involves grouping together jobs which show homogeneity of work behaviors. If the degree of homogeneity is high and if test/ability linkages so permit, these occupational groupings can be subsumed under the same test battery. Current research and development efforts include Law Enforcement Occupations, Scientific Aide and Technician Occupations, Apprentice Occupations, and Professional and Administrative Occupations (these comprise 100+ occupations and have been subsumed into six occupational groupings).

"Coupled with the streamlining of test batteries, described above, this grouping of jobs into occupational bands, is gradually instituting a testing system which is optimally efficient and effective, in terms of cost, administrative and operational viability, and psychometric integrity" (U.S. OPM, OPRD, correspondence, 3/3/89).

CONVERSION OF VERBAL REASONING TESTS TO A LOGIC-BASED MEASUREMENT SYSTEM

OPM's work on logic-based measurement is very recent and has been published in *Personnel Psychology* (Colberg, 1985). An OPM paper on the work is also available for interested readers (Colberg & Nester, 1987). An OPM, OPRD, report on this work follows.

"Logic-Based Measurement (LBM) was instituted recently in OPM in one of its new examinations, that for Contract Specialist. Also, the new set of examinations for professional and administrative occupations is being developed according to the principles of LBM.

"LBM is a technology of measurement in which all measures (i.e., test items) are constructed according to the inferential formulae of logic. Adherence to these formulae makes possible for all verbal reasoning tests the precision of measurement that had been available heretofore only for mathematical reasoning tests. In other words, LBM insures that each test item in a verbal reasoning test (such as reading comprehension or inference) conforms to a logical formula that is objectively and necessarily valid. LBM thus brings into verbal reasoning tests the inferential objectivity of mathematical reasoning tests.

"Additionally, the LBM technology has various other psychometric advantages. First, in defining the reasoning ability constructs measured by tests with the precision of logic, LBM makes possible the removal from the spectrum of measurement of redundant ability constructs and of redundant, cost-ineffective marker tests. Second, through its use of taxonomies based on the formulae of logic, LBM serves to insure that every operational test question measures the intended construct and that the sampling of reasoning processes is sufficiently broad.

"Thus, in the context of employment testing, LBM serves various fundamental purposes such as improving precision of measurement, decreasing litigation

vulnerability and increasing cost-effectiveness (through the removal of redundant measurement from tests).

"Thus far, OPM has used LBM only in the context of tests developed under the classical test theory model. But, it is expected that LBM will lend itself particularly well to a Latent Trait Theory Model because LBM enhances the possibility of achieving unidimensionality and definitive linkage between item and construct, both of these conditions being fundamental assumptions in a Latent Trait context" (OPM, OPRD, correspondence, 3/3/89).

WORK FORCE QUALITY ASSESSMENT RESEARCH

"Work force quality is a critical concern facing every employer, including the federal government. Changing demographics and educational deficiencies will have a clear impact on the nation's labor market and on our selection policies in the coming years. Some argue that the impact of these changes on quality is already being felt.

"To address these concerns, the Office of Personnel Management has put into place a comprehensive research program to assess the competence of its applicants, new hires, long-term workers, and people who leave. The program will collect quality indicator information (such as scholastic and academic grades, level of education, amount of experience, and specialized training) for applicants and analyze trends evident over the years. Job performance will be measured using research supervisory ratings for the incumbent work force.

"Essentially, OPM is expanding upon test validation strategies, and incorporating new measurement designs in all of its recruitment, selection and retention programs. These programs will be evaluated more systematically than ever before" (U.S. OPM, OPRD, correspondence, 3/3/89).

SUMMARY AND RECOMMENDATIONS

The U.S. Merit System Protection Board is conducting independent evaluations of the effectiveness of the federal merit system. These evaluations have identified a variety of problems as well as successes, which were noted in the above article. The U.S. Office of Personnel Management (OPM) has taken many steps to address problems in personnel assessment, and continues to be a primary contributor of public sector assessment innovations.

In the opinion of the author, there have been many problems for federal and other public sector assessment professionals. These problems may be some of the causes of the inadequacies identified by the MSPB. Working to improve the fairness and quality of tests with reduced resources in an atmosphere of negativism towards government employees was a significant problem in most of the 1980s. Some examples of such problems include: the philosophy of the 1980s ("doing more with less") in government was not viewed as particularly supportive of government workers and was not attractive to job applicants; the lack of support for federal personnel improvement programs (such as the Intergovernmental Personnel Act covered later in this report), elimination of most Regional Psychologist positions in OPM, the reduction of OPM's Personnel Research and Development Center (PRDC) staff by about 50%, including the elimination of the staff which worked to help state and local governments make assessment improvements (State and Local Section, PRDC); the identification of government regulation as interference and as harmful to the economy; and the lack of significant pay increases for government managers and employees at the working levels.

Another problem has been the lack of clear and consistent policy direction within the federal government on critical assessment issues. This is illustrated by the years of interagency efforts it took to produce the Uniform Guidelines on Employee Selection Procedures; the lack of progress in obtaining interagency concurrence to update those guidelines; the years of lack of progress on the development of replacement procedures for the PACE examination; and differences in the use of tests between federal agencies.

As an example of the last item, the U.S. Department of Labor is using a controversial "within-group" applicant referral procedure with the General Aptitude Test Battery. Other federal agencies practice different referral procedures.

The 1989 Volcher Commission Report on "Rebuilding the Public Service" addresses the "quiet crisis" in government of too many of the best senior executives leaving the public service and not enough young talented people who are willing to join the public service. A key recommendation of this report is to strengthen OPM. The report identifies many of the problems OPM faces including: a huge clientele, limited staff and funding, budget cuts in recent years, turnover at the top of the agency, and fallen staff morale (National Commission on the Public Service, 1989, pp. 39-40).

Another trend at OPM, which is related to the resource reduction trend, has been a decrease in the helping and leadership role of OPM in public sector assessment. In the 1970s and very early 1980s, OPM was the major source of information on research and innovation in public sector assessment. During that time, OPM published a variety of reports and guides on such topics as legal developments, research summaries of existing literature, guides on the development and use of various assessment techniques, as well as a variety of complete original research reports. These helpful reports and literature were widely distributed, read, and used to upgrade public sector assessment practices. In addition, OPM offered technical assistance and training on public sector assessment topics through central and regional offices. Since the early 1980s, the information from OPM has decreased dramatically. The information submitted for this report indicates that OPM may be moving to reclaim its former role as a major resource in public sector assessment innovation and knowledge contribution.

Despite the problems noted earlier, OPM continues to make noteworthy contributions to personnel assessment improvement. OPM innovations include efforts to: group jobs for examining; simplify tests; reduce adverse impact; develop job-specific ability tests; develop a technology of logic-based measurement; develop a research program to assess the quality of the work force and determine trends in work force quality. Additional OPM innovations are

summarized in other sections of this report, such as: innovative biodata instrument development and research, the development of procedures for testing the disabled and making test accommodations, the development of the behavioral consistency E&T method, and the development of meta-analysis technology to summarize and evaluate personnel research findings.

The National Commission of Testing and Public Policy should address: 1) the policy issue of support for assessment improvement efforts, and 2) the direction which those efforts should take.

Specific support recommended

- Restoration of Intergovernmental Personnel Act funding for assessment improvement efforts and innovations to the level of staff and funding provided in the 1970s, including the re-establishment of regional offices to provide technical assistance and training, and encourage innovations throughout the U.S.

- Restoration of the Office of Personnel Research and Development of the U.S. Office of Personnel Management to 1970s level of staffing, funding, research, and technical assistance and training in assessment to federal agencies and state and local governments. The state and local section of OPRD should be re-established.

Policy direction recommended

- Encourage research and innovation in assessment, such as that illustrated in the various sections of this entire report. Such efforts need to become the norm in assessment, rather than the efforts of only some employers.

- A clear consistent policy direction on the use of tests and test scores in federal agencies.

As illustrated in many sections of this paper, it takes resources and an interest in and commitment to fairness and innovation, to develop and carry out assessment programs which reduce group differences and maintain

101

validity. Traditional ability tests and traditional paper and pencil tests have higher adverse impact than the alternatives reviewed in this paper. As compared to innovative methods, traditional paper and pencil tests can be purchased or developed, administered and scored with considerably less resources, and perhaps with less expertise on the part of the developer. Perhaps that accounts for the heavy use of paper and pencil tests in educational admissions and employment testing.

APPLICATION OF TECHNOLOGY TO ASSESSMENT

USE OF VIDEO

Background Information

A major trend in public sector assessment is the development of job-specific tests. Video technology is being used to develop realistic job previews and to develop and conduct work simulation tests.

Such tests have been designed to assess a variety of requirements for different occupations. One may question whether the use of video tells us anything more about a candidate's ability to do a job than a paper and pencil test format (R. Reilly, correspondence, 5/2/89). Video simulations can be used to evaluate requirements which cannot be assessed as readily with strictly paper and pencil devices, such as: listening ability, visual observation ability, verbal language translation ability, note-taking ability, and application of job knowledge.

One major advantage of videotaped stimuli in assessment is the standardization of the task presented to examinees. In addition, the problem of obtaining role players, actors, or others needed for a simulation is simplified. Such persons are only needed to produce the video; they do not need to be present for each candidate. In addition, use of raters can be more efficient with video. This can increase reliability and validity (Quero, correspondence, 5/89). Videotaped tests and candidate responses may be appropriate alternative selection procedures. Such tests may have less adverse impact.

Following are examples of the use of video in public sector assessment. The examples include: a videotaped realistic job preview; videotaped tests or test segments for Court Interpreter, Parole Hearing Officer, Correction Officer, Corrections Lieutenant, and Postal Inspector; and videotapes for orientation of candidates and training of assessors.

Research on Realistic Job Previews (RJP)

Providing job applicants with complete and accurate information on a job can help to reduce turnover problems. It can also help to accomplish affirmative action goals.

It is very important that a job preview be an accurate representation of the job. Overly negative presentations can discourage candidates from applying and can create adverse impact. In one case, overemphasis of the physical demands of a job discouraged many females from applying. This resulted in poor work force representation of females (R. Reilly, correspondence, 5/2/89).

Templer & Tollever present information on the value of RJPs as follows:

"Unmet expectations lead to job dissatisfaction and consequently high job turnover or lowered performance. ...a number of experiments were designed to present a more balanced view of jobs to possible recruits.... In some cases the true picture was presented in written form, sometimes in film format, and occasionally orally. While these experiments dealt primarily with lower level entry positions, evidence led researchers to conclude that a realistic job preview, as predicted, lowers turnover and, additionally, does not reduce the potential recruitment pool. RJPs may therefore, be a partial answer to problems of target group retention in managerial jobs. Expectations can be made more explicit and realistic without reducing recruitment pool numbers" (Templer & Tollever, 1983, pp. 213-214).

Tomes and Davey (1987) of the state of South Carolina reported on the development and use of a videotaped realistic job preview for the job of Corrections Officer Trainee. In their presentation they reviewed theories on the effects of RJPs. RJPs may "vaccinate" the applicant to be ready for the job; the applicant will not be "turned off" by the job. RJPs may enhance the organizational match between the desires of candidates and employers, and result in the selection of candidates who have a higher personal commitment to the organization. With RJPs, applicants may view the organization as more honest and caring. The applicant may

have more realistic expectations; and may view the job as a challenge (e.g., "This job is difficult but I can do it").

Tomes & Davey suggested that RJPs may be most appropriate for entry-level jobs where there is a low selection ratio (i.e. many or most applicants are hired), and where there is low unemployment. In these situations, applicants have a lot of choice concerning the job they accept.

A 25-minute videotape "Fear is a Good Sign" was prepared in Correctional Institutions in South Carolina. The tape provides a view of the typical duties of Corrections Officers. In the tape, Corrections Officers comment about their jobs, the problems, challenges and rewards. Candidates see officers dealing with inmates, and hear and see the typical environment in which they will be expected to work, behind closed bars.

The tape has been used on an experimental basis in South Carolina. One-half of all Corrections Officer Trainee job applicants are shown the video. All applicants complete a job expectations questionnaire. South Carolina is conducting research to determine if use of the RJP reduces turnover, or relates to other criteria such as job satisfaction. Turnover is a serious problem in these positions in South Carolina. Twenty-five percent of new hires leave employment within six months of hire.

EXAMPLES OF VIDEO USE

Interpreter Testing. The New York State Office of Court Administration employs interpreters in more than 50 languages, and has 120 Spanish Court Interpreters. This agency has developed videotaped material which is used in the assessment of applicant bilingualism, oral interpretation and other skills.

The major tasks of interpreters were defined in a job analysis study of the positions (Minter, 1988). For example, two of the five major job tasks are:

— Oral courtroom interpreting, e.g., translates verbatim questions into Spanish for the witness and the witness' answers into English for the attorneys, jury, and judge.

- Translating written material and audio tapes, e.g., translate from English to Spanish documents such as charges, waivers or extradition and reads these to defendants. Translates audio tapes of wiretaps of individuals speaking in Spanish.

Part of the selection procedure developed to assess the job requirements derived from the job analysis was a standardized work simulation.

"Scripts were developed based on actual cases in Civil, Family, and Criminal Courts. In each case there was an English speaking attorney and a Spanish speaking witness. Candidates were required to translate into Spanish everything spoken in English and vice versa.... A videotape of actors reading the exam script was made. Each candidate was played the tape on a television screen, and simultaneously an audio tape of his/her oral interpretation was made. ...candidates were given two short written passages (one in English and one in Spanish) to review for five minutes and then a sight translation of the passages was included at the end of their audio test tapes" (Minter, 1988, pp. 7-8).

A group of bilingual raters were trained to evaluate the tapes made of candidate responses to the above test situations. Two raters evaluated each candidate. Rating scales were developed for six rating factors such as: "Accuracy/ Comprehension from English to Spanish," and "Clarity/Pronunciation in Spanish."

The New York State Office of Court Administration evaluated this assessment procedure and determined that it worked well. It allowed for a standardized task for all candidates. The use of videotape received positive feedback from candidates and facilitated the scheduling of candidates and raters. There was considerable cost savings compared to the previous procedure of having live actors for each candidate. The court office plans to use the videotape testing method in other languages in the future (Minter, 1988, p. 11).

Parole Hearing Officer. The use of video in an "Oral/Work Simulation Test" of the Pennsylvania State Civil Service Commission was described as an "...alternative

method for selecting individuals for administrative types of positions when a more extensive testing instrument than a traditional oral is warranted but time and resources do not permit full scale assessment centers to be used" (H. Sheibley, 1986, p. 1).

Sheibley conducted a job analysis of the Parole Hearing Officer positions. Employees have responsibility for conducting all types of parole hearings and parole interviews. For example: they make determinations and recommendations in accordance with laws and policies governing disposition of cases. They review arrest records, progress and conduct reports, classification summaries, parole summaries, and psychological and psychiatric reports. They hear cases. They recommend in writing to the Parole Board if an inmate should be released on parole, or if a parolee should have their parole continued or revoked (H. Sheibley, 1986, p. 5).

A two-part examination was developed to simulate the most important aspects of the job. The first part of the examination was the Oral/Work Simulation Test. In this test part, applicants were first given 30 minutes to review information needed to write a hearing report "General Rules Governing Hearings." The applicants then observed a videotape of a simulated Parole Hearing. Applicants were permitted to take notes while observing the tape. Applicants were then required to write a hearing report based on the information they analyzed and their observations and notes. Applicants can refer to the document provided earlier when they prepare their report.

The hearing reports of applicants are rated by a panel of trained raters using rating scales developed for this purpose. The factors assessed are: thoroughness and accuracy, organization and clarity, logic and sound decision-making, and written communications.

The second part of the examination was a structured oral examination with situational questions representing the types of situations a Parole Hearing Officer might encounter soon after being hired.

Corrections Officer Observational and Writing Skills Assessment. G. H. Friedlander of the state of New Jersey Department of Personnel has developed a series of

videotaped vignettes to assess the observational and writing skills of entry level corrections officer candidates in state and local governments.

Observing incidents, remembering facts, and writing accurate narrative incident reports are critical tasks of corrections officers. Friedlander (1988, p. 1) described an example of the outcome of a poorly written corrections officer report. An inmate had a series of incidents of violent behavior in prison but wasn't "locked up" because written reports of the incidents were defective and did not "hold up" in disciplinary hearings. The inmate was allowed to stay in the general prison population where he later killed Captain Felix Mokechick.

The video consists of a practice exercise and three prison incidents which examinees must observe, and then describe in writing what happened (e.g., persons involved, times, location, objects used, etc.). Each vignette is administered separately. Total administration time takes about 45 minutes. Candidate products are evaluated independently by two raters using written criteria.

The test was tried out and differentiated between candidates. Data on the test revealed no differences in writing ability of candidates by race or sex (Friedlander, 1988, p. 4).

The state of New Jersey has also developed videotaped tests or test segments for Police Guard, Hearings Officer – Inmate Disciplinary Program, Corrections Lieutenant, and Corrections Captain. The Corrections Lieutenant test includes a videotape test on supervisory problems, as well as a 22-minute video inmate count-taking exercise (G. H. Friedlander, personal correspondence, 3/3/89).

Corrections Lieutenant. Paul Kaiser (1988) of the New York State Department of Civil Service developed a four-part promotion examination for the evaluation of Correction Lieutenant candidates. The parts were:

1. A memory test to evaluate the candidate's knowledge of the rules, regulations and department directives which needed to be known from memory.

2. An open-book test to evaluate knowledges that were critical but for which the candidate typically would be able to refer to references on the job.

3. A video test designed to present nonwritten material which would evaluate skills and abilities that could not be measured in other components. The candidates observed six video scenes and then answered questions in a test booklet based upon what they observed.

4. An incident simulation test presented candidates with an emergency situation, a stabbing investigation situation, a supervisory problem, and a series of "day in the life" situations which incumbents have to deal with frequently.

One hypothesis was that the video-based test would have smaller group differences than the other test parts. The mean score differences of the 744 candidates were in fact smallest for blacks on this part of the test. This test part also had the lowest correlations with the other test parts, indicating that it was in fact measuring different requirements. As stated by Kaiser (1988): "...since videos may not have the verbal loading that is found in written tests, they serve to minimize adverse impact..."

Assessment Center Exercises. The U.S. Postal Service uses assessment center exercises as one part of the selection process for Postal Inspector positions. One segment of the assessment center evaluates candidates on their ability to extract and interpret information and write a report on the information obtained. Candidates observe and listen to videotapes of situations encountered by Postal Inspectors and write reports on what they observed and heard. For example, one video scene shows the hold up of a Postal Clerk, another shows a Postal Inspector interviewing a bank official on a credit card cash withdrawal transaction by a customer who had obtained and illegally used multiple credit cards. Video tapes are also used by the Postal Service to help train assessors (Ocasio, 1988).

Examples of Video Use for Orientation, Training and Other Purposes

U.S. Office of Personnel Management. "Videotapes have proven to be a valuable tool in the training of individuals who plan to serve on panel interviews or as assessors for an assessment center. The video is designed to facilitate objectivity in assessing observed behavior. The video has been used in several programs, such as the Professional and Administrative Career entry for selected occupations, the Presidential Management Intern Program, and the Administrative Law Judge examination, when interviews are used in the examination. The video plays a vital role in training individuals how to observe, record, categorize, and rate observed candidate behavior. It permits the trainees to practice with a standardized simulation exercise. All of the videotapes are accompanied by training manuals to ensure that they are being properly used. It is very transportable and can be used without a trainer, when the situation is not feasible to conduct a formal training class" (U.S. Office of Personnel Management, Office of Personnel Research and Development, correspondence, 3/3/89).

Video on Assessment Centers. Dennis and Sherry Joiner (1985) have developed and utilized a 50-minute videotape on "Assessment Centers: What Are They?"

This tape is used to provide orientation to candidates, as well as potential assessors, about the assessment center process.

> "Candidate orientation is a critical requirement for the success of any assessment center examination process. ...It is important to...equalize the knowledge of the assessment center process within the candidate group. This will reduce any possible unfair advantage or perception that someone had and unfair advantage due to prior exposure or different levels of familiarity with the technology" (Joiner, 1985 p. 172).

The videotape viewer follows a group of six candidates through an orientation session. Four common job simulation exercises are shown: a leaderless group discussion, an in-basket exercise with an interview, an oral presentation,

110

and two examples of a role play or subordinate counseling exercise. Each exercise includes commentary explaining the rationale for the exercise and other orientation information.

South Carolina. Video is used for candidate orientation and instruction by the state of South Carolina. Videotapes provide instructions to candidates taking written tests. For example, magnified examples of how to mark an answer sheet and complete other required forms are shown to candidates.

The state of South Carolina has used closed-circuit TV to administer a job analysis task inventory to 2,000 employees in a 24 hour period (Mayes, 1980).

Summary and Recommendations

Video technology is being used in public sector assessment in such areas as realistic job previews, candidate orientation, training of assessors and raters, and in work simulation exercises for a wide variety of job requirements and occupations. A number of studies have shown high applicant acceptance and low or no group differences in work simulations using video technology.

The National Commission on Testing and Public Policy should recommend expanded application of video technology to personnel assessment. Continued research on the cost effectiveness, relative validity, and adverse impact of this technology should be encouraged.

USE OF COMPUTERS

Background Information

A survey of computer use in a sample of merit systems illustrates the heavy reliance placed on computer technology to conduct the day to day business of public personnel assessment. Of 51 responding jurisdictions, 78% used mainframe computers, 41% used minicomputers, and 69% used personal computers (Eisenhart, 1986, p. 3). As the percentages indicate, many jurisdictions use a number of types of computers. The survey found the following to be the most common uses of computers in the responding merit systems: exam scheduling of candidates, administrative tracking of candidates, written exam scoring, eligible list establishment and maintenance, certification of candidates, and item analysis (1986, p. 2).

In an article on "Computer Applications to Personnel..." (Darany, 1984) defined areas of savings from use of computers as cost-effective productivity aids. Darany's categorizations, as well as other references, identified these areas of current use of computers in assessment:

- word processing,
- applicant tracking,
- job analysis,
- test analysis,
- item banking,
- test administration,
- training,
- personnel information management, and
- candidate/employee guidance systems.

Word Processing

As in the private sector, the primary use of word processing in public sector assessment has been to increase the accuracy and speed of document creation and revision. Such features as spelling check, automatic justification, word-wrap, and the need to only proofread newly revised sections of documents, help to obtain high quality, readable, error-

free test materials. Darany estimates that "...we may be able to save 25-200 percent of the work involved in creating a final version of materials traditionally typed in a personnel office, if an effective word processing system is used instead of a high quality electric typewriter" (1984, p. 454).

Applicant Tracking

Use of computers for applicant tracking in the city of Denver, Colorado, helped to solve problems of speed and accuracy in such areas as: recruitment, scheduling candidates for tests, scoring tests, notifying candidates of test results, maintaining records on candidates, and certifying candidates to city agencies for employment consideration. In 1982 Denver purchased and installed a commercial mini-computer with a software package called "TRAC." This system was designed for government merit systems and had proven effective in other places. To obtain the system, McClung "traded" funds from four of his budgeted clerical positions. He "...was able to show that TRAC would more than pay for itself within the five-year lease/purchase period" (McClung, 1985, p. 31).

McClung reported that this system has been extremely successful in Denver. Workload has gone up while the staff size decreased. The mini-computer was key to survival in this situation. The computer is also being used in other applications in Denver, such as salary survey processing, and fringe benefit studies.

As stated by Darany (1984, p. 455):

"A computer-based test applicant tracking system should allow for productivity gains in a range of 25 to 100 percent. It should also produce a more dependable time frame within which the testing activities take place. Finally, it may develop information useful to assess the quality of one's tests and the adequacy of current lists."

Computerized applicant tracking systems facilitate the conduct of adverse impact studies to identify occupations where selection processes may be unfairly screening out minority group members. In addition, such systems

113

free test materials. Darany estimates that "...we may be able to save 25-200 percent of the work involved in creating a final version of materials traditionally typed in a personnel office, if an effective word processing system is used instead of a high quality electric typewriter" (1984, p. 454).

Applicant Tracking

Use of computers for applicant tracking in the city of Denver, Colorado, helped to solve problems of speed and accuracy in such areas as: recruitment, scheduling candidates for tests, scoring tests, notifying candidates of test results, maintaining records on candidates, and certifying candidates to city agencies for employment consideration. In 1982 Denver purchased and installed a commercial mini-computer with a software package called "TRAC." This system was designed for government merit systems and had proven effective in other places. To obtain the system, McClung "traded" funds from four of his budgeted clerical positions. He "...was able to show that TRAC would more than pay for itself within the five-year lease/purchase period" (McClung, 1985, p. 31).

McClung reported that this system has been extremely successful in Denver. Workload has gone up while the staff size decreased. The mini-computer was key to survival in this situation. The computer is also being used in other applications in Denver, such as salary survey processing, and fringe benefit studies.

As stated by Darany (1984, p. 455):

"A computer-based test applicant tracking system should allow for productivity gains in a range of 25 to 100 percent. It should also produce a more dependable time frame within which the testing activities take place. Finally, it may develop information useful to assess the quality of one's tests and the adequacy of current lists."

Computerized applicant tracking systems facilitate the conduct of adverse impact studies to identify occupations where selection processes may be unfairly screening out minority group members. In addition, such systems

113

facilitate analysis down to the test item level. Separate item analysis is conducted by race to identify test questions which may be unfair. Such test items are then reviewed to make certain they are essential. Test items are retained, replaced or revised based on the results of the review and the availability of alternative items. Such use of item analysis data in applicant tracking systems is occurring in public sector jurisdictions which have the necessary computer equipment and software, the necessary staff resources, and sufficient applicant volume to support such analysis.

Job Analysis

A symposia on "Computer-based Job Analysis" (Trattner et al., 1980, pp. 13-14) illustrates the variety of uses of computer technology in job analysis. Marvin Trattner described the use of CODAP (Comprehensive Occupational Data Analysis Program) in the federal service. Robert Lockwood described use of SPSS (Statistical Package for the Social Sciences) to analyze a task inventory for 31 job titles and prepare a set of test specifications. Phyllis Mayes of the state of South Carolina described a study of 10,000 clerical positions based on a sample of 2,000 employees who were administered a task inventory using closed-circuit television. The South Carolina data was analyzed by computer using a cluster analysis procedure. A study of technology transfer by the National League of Cities was also reported on by Michael Frank.

The National League of Cities obtained a federal grant to fund a study of computer assisted human resource management systems in the areas of classification, performance evaluation, productivity measurement, selection, and training. The study investigated human resource management systems in use at the time and collected data on the informational needs of ten cities throughout the U.S. The findings of this study:

"...indicate that computer applications are limited mainly to record processing and CODAP is the only successful technology transfer at present. Problems encountered...include the modification of technology to novel settings, the disappearance of technology when it

114

is not subsidized with grants and the like, and the fact that policy leaders are often ill-equipped to understand and deal with new technologies..." (Trattner, et al., 1980, p. 14).

Marlene Goodison, who conducted the study for the National League of Cities, reported CODAP to be: "The most extensive human resources management system on the scene. ...Some cites are already using CODAP, some are in the process of implementing it, and others are contemplating its use" (Goodison, 1980, p. 4).

CODAP consists of a set of computer programs used to analyze job content data. Job data is collected by preparing an inventory of tasks which is administered to large numbers of incumbents or supervisors. Data is collected on such questions as: Is each task performed or not? What is the relative amount of time spent on each task? CODAP is used to analyze the data. The system was developed by the U.S. Air Force and is widely used in the military as well as in some federal, state, and local governmental agencies. The results of a CODAP analysis can be used to help classify jobs, prepare accurate job descriptions, develop training programs, design selection procedures, develop performance standards, and for other purposes.

In a presentation on "Automating the Examination Process in the 80s," use of CODAP by the classification and examination sections of San Bernadino County, California, was described. The study of 2,400 clerical positions produced data useful for many purposes (Darany, 1980, p. 34).

The Salt River Project developed a computer-based job evaluation system to be used in evaluating the worth of jobs. The system relies on use of UCLA's BMDP (Biomedical Computer Programs) to analyze data. The purpose of the study was to design an objective nondiscriminatory job evaluation technique (Horney & Dertien, 1981, pp. 19-20).

The above studies illustrate the wide range of uses of computer technology in job analysis and job evaluation in the public sector. This technology allows more sophisticated data analysis than was previously practical within many public jurisdictions. When properly conducted, such

analysis allows more accurate definition of jobs and more accurate determination of job requirements.

Test Analysis

The statistical software packages mentioned above (e.g., SPSS, BMDP) are used in some public sector jurisdictions to perform a variety of test analysis and examination improvement studies, such as: reliability studies, criterion-related validation studies, test fairness studies, and adverse impact studies. One type of such analysis, the statistical analysis of test items to help reduce group differences in test performance while maintaining necessary test content coverage, was described earlier under "Applicant Tracking." A number of jurisdictions have developed their own tailor-made software to do such analysis or have purchased consulting services or other software packages for such purposes.

The U.S. Office of Personnel Management recently obtained a sophisticated state-of-the-art computer package called LISREL for structural equation modeling. This tool will allow the study of networks of variables. Researchers in OPM have previously used this type of analysis in personnel assessment studies of some military jobs and will be aided considerably by this new tool (U.S. OPM, 1989).

Another example of use of computers to assist in improving public sector tests is in the area of readability analysis. Some jurisdictions use readability analysis software to help assure that the language level of written tests, job announcements and other documents are not above the reading level required for the positions.

Item Banking

Test questions (items) represent a major resource of assessment professionals.

"Computer technology provides a means of maximizing the utility and quality of that resource. Items (questions) may be categorized according to subject matter content. They may then be stored in a computer system, along with data regarding history of use and item quality. As items are used in tests, new history data are

added. Consequently, an item bank reduces test development time, improves the efficiency of professional staff, and produces higher quality written tests. It is not necessary to produce all new items for every testing need. Previously used items can be selected from the appropriate subject matter categories. Within a category, items can be selected on the basis of their past performance. And items can be improved continually through revision based upon past performance" (Darany, 1984, p. 457).

Western Region Item Bank (WRIB). Darany and French described the development of the Western Region Item Bank (WRIB). "WRIB is based on a conceptual item bank model developed by T. Darany in Michigan in 1970 that was successfully implemented in 1975 in a single jurisdiction by the State of Missouri" (Darany & French, 1986, p. 1). Work on the bank grew out of discussions of common problems of members of the Western Region Intergovernmental Personnel Assessment Consortium (WRIPAC) in 1979. Few small sized jurisdictions could afford the equipment and programming costs to develop and operate an item bank. This led WRIPAC to develop a cooperative item bank which is housed in and managed by San Bernadino County, California. The bank has been operational since 1981.

WRIB is the first cooperative multi-jurisdictional test item bank. This bank had 99 members in 18 states as of June 1988. Although the bank primarily services West-coast jurisdictions, membership includes some Midwest and East-coast states such as Ohio, Wisconsin, Michigan, Vermont and Rhode Island.

Members of WRIB pay a yearly membership fee of $1,500. They agree to: test security provisions in a contract for services; provide 80 hours of staff time per year to review and edit test questions for the bank; and agree to submit items for entry into the bank or to review, edit, and classify items provided by other members.

The bank contains over 30,000 test questions and 117,000 item histories. Users report fast turn-around time and satisfaction with the service. Services include:

obtaining test items for review, preparation of master test booklets, test scoring, statistical analysis, and reports on tests and test items. Items are available based on a variety of characteristics including: content, format, difficulty, and discrimination index.

The success of WRIB has led other consortia, such as the Region Six Personnel Assessment Council (RESPAC), to develop a similar cooperative item bank now operated by the state of Oklahoma (Darany & French, 1986, p. 6).

Other Public Sector Item Banks. The states of New York, Maryland, New Jersey, and Pennsylvania, and the City of Baltimore, Maryland, have developed computerized test item banks. The California State University System has a system called SOCRATES. Summary information on the New Jersey bank and the New York bank follows.

The New Jersey State Department of Personnel has had an "Automated Test Generation" (ATG) system in use since 1985. The system stores oral, performance, essay, and multiple-choice questions on a Prime supermini computer. The system can link job analysis data (tasks and knowledge, skill and ability requirements) to test items. The bank contains over 50,000 records and services 75 users at three sites. This system was developed under contract beginning in 1983 with Assessment Systems, Inc. of Philadelphia (Kaido, 1988, p. 5).

Beginning in 1985, the New York State Department of Civil Service began development of a Test Item Data Bank (TIDB). The system is now operational. It is housed on a UNISYS mainframe and will have 90 users assigned to 20 terminals. Over 50,000 multiple-choice written test items are stored on the bank. The bank stores items; item history information, such as when used, and the names and dates of the tests in which used; item difficulty and discrimination statistics; appeal history; author; and content category. Items can be accessed by one or more of these criteria (Kaido, 1988, p. 5).

In summary, computerized test item banks have been developed in a number of public jurisdictions and regional consortia to improve the quality and cost-effectiveness of tests and the test development process.

Commercially Available Micro-computer Item Bank Software. The item banks reviewed above are all designed for mainframe computers or minicomputers. The hardware cost and the cost of development of the software for some of these item banks may be beyond what many public sector merit examining agencies can afford (Quero, correspondence, 5/89). Some of the item banks described earlier have cost from 1/2 to 1 million dollars to develop. Others are considerably less expensive but are still in the five figure range.

The cost of an item bank has recently dropped dramatically as a result of the availability of powerful microcomputers with large storage capacities and high processing speeds. Such machines can now be purchased for a few thousand dollars. Item bank software is now available at reasonable cost for such machines.

Three available microcomputer item bank software systems are: STAT–Standard Testing and Analysis for Teachers (available from Applied Measurement Professionals, $495), PAR (Economics Research, Inc., $1,000), and MICROCAT (Assessment Systems Corp., $3,100). The MICROCAT system can be used for classical test development, item response theory and adaptive testing. Each of these packages are user friendly and cost effective (L. Quero, correspondence & enclosures, and vendor catalogues, 5/89).

Test Administration

Adaptive Testing. The most extensive research on the application of computer technology to applicant assessment has been by the military. The Office of Naval Research and the Army Research Institute have developed a technology known as computerized adaptive testing (CAT). A computerized bank of test questions and sophisticated computer software are used to interactively assemble and administer a test to an applicant. Test questions are selected by a computer program based upon the applicant's response to previous questions. The objective is to obtain a very accurate assessment of the applicant's true ability in a short period of time, with the test being tailored to the ability level of the

applicant. This technology is being applied to the Armed Services Vocational Aptitude Battery, which is used to assess applicants for entry into the military services. It is also in use with Architect licensing examinations, and is in the developmental stages for Nurse licensure testing.

Work Simulation Testing. Another sophisticated technology is computerized simulation testing. This approach uses the branching technique described in the article on latent image testing in the work simulation section of this paper. Such a test has been developed by the National Board of Medical Examiners for the training of physicians.

Students are presented with information about patients. The prospective physicians are able to ask questions of the patients and obtain information, order tests, review test results (e.g., view X-rays, review blood test and electrocardiogram results, hear a patient's heartbeat), and prescribe treatment programs for patients. Depending on their decisions, the patients may survive or expire along with the student's grade. For such life and death decisions, computer simulations have considerable advantages, especially for the patients. This technology has also been used in flight simulator training and in war games in the military.

The scoring of the physician simulation test goes considerably beyond the typical correct/incorrect scoring procedure and will include such factors as: proper diagnosis, proper treatment, cost of the treatment, and evaluation of the risk of the treatment for the patient.

Examples of Merit System Uses. The results of a 1986 survey of computer use in merit system selection indicated that much more has been written about computer administered tests in merit system agencies than has been achieved (Eisenhart, 1986). However, there are some recent examples of successful innovations.

One of the most common procedures in a few merit system agencies is to use a computer to administer a test. For example the state of Maryland tests applicants for driver's licenses on their knowledge of driving rules and regulations by computer. Computer administered and scored typing tests, word processing tests, and data entry tests have also been developed. An example of such a procedure for a

clerical job is noted in the section of this paper on work simulation tests.

In a symposium on computer assisted testing in government jobs, the Advanced Research Resources Organization described cognitive and perceptual-motor, computer-assisted tests being developed for the military (Myers et al., 1983, p. 75). A later symposium on the same topic described the advantages of computer-administered tests, a basic language program to develop and automate simulation tests, and a computer-administered interest inventory (Darany, Jacobsen & Davey, 1986, pp. 135-148).

The New York State Department of Civil Service has developed and used a computer administered work simulation test for the job of Motor Vehicle Representative. Employees in these jobs provide direct service to the public. The jobs have a high turnover rate, especially in New York City. Employees generally work at a counter and deal with long lines of customers. They record licensing and registration transactions on a computer terminal. There is considerable pressure to work quickly and accurately.

A test was developed which is administered at the job site on the computer terminals used by employees in these positions. By conducting the test at the actual work site, applicants receive a type of realistic job preview. The test had five subparts: coding, arithmetic, reading comprehension, proofreading, and keyboard accuracy. The test parts were derived from job analysis results. Questions were simple and simulated job tasks. The test was speeded to simulate the pressure of the job. Applicants were sent a description of the test and information on how to operate the terminal prior to taking the test.

The test was first given in 1985 to 714 candidates. Over half the candidates were Black or Hispanic. A higher proportion of non-white candidates were in hirable scoring range on the computer-assisted examination than with the previous written test of 1984. A questionnaire sent to supervisors found a "...high level of satisfaction with employees hired from the lists resulting from the examination.... Supervisors clearly indicated that the content measured on the

computer-assisted examination was job-related" (New York State Department of Civil Service, 1987, p. 9).

Training

Darany reported on the use of a computer-administered training program in typing skill. A computer-based learning laboratory was established at San Bernadino County.

"A fairly consistent 25 wpm improvement in typing speed was found after 10 hours of intensive computer assisted instruction. This was true both for persons of modest skills (e.g., starting with 25 wpm) as well as higher skills (e.g., 65 wpm). The limited amount of evidence of automation for training purposes suggests that well-designed computer-assisted instruction may be very effective in the areas of skill development, orientation, and skill refreshing training..." (Darany, 1984, p. 458).

Personnel Information Management

A variety of computer software is available to manage information. For example, Darany reported on the potential use of spreadsheet and data base automation tools for budget development and forecasting, tracking unemployment insurance claims and awards, and evaluating the success of recruitment sources (Darany, 1984, pp. 458-460).

An article on the initial use of microcomputers in the Pennsylvania State Civil Service Commission illustrates the use of such spreadsheet and database microcomputer software. Lotus 1•2•3 was used to analyze job analysis data and adverse impact data. dBase III+ software and R&R Relational Report Writer were used to plan and organize the examination and recruitment schedule of this merit system which recruits and tests over 100,000 candidates for over 800 job classes per year (Sproule, 1988, p. 11).

A computer-assisted support system for staffing decisions was developed in the U.S. Office of Personnel Management. The system is available on diskette and will operate on an IBM PC or a compatible machine. The system

helps one to choose the best personnel selection measures for a given occupation. The information which the computer program provides is based upon thousands of validity studies. Some of the information provided includes: an estimate of the validity of a proposed selection system, an estimate of the economic value of a given screening measure, and information on the expected adverse impact of several measures (McDaniel, 1987). A description of this computer-assisted Staffing Decision Support System (SDSS) was published in the spring 1989 issue of *Public Personnel Management* (McDaniel & Schmidt, 1989).

Candidate/Employee Guidance Systems

The U.S. Office of Personnel Management (OPM) is developing a computer-based career guidance and occupational information system for the federal government. The system is being designed to help in the recruitment process by providing occupational information to applicants, such as: descriptions of federal jobs, minimum qualifications requirements for jobs, salary information, major employing agencies, and geographic locations of jobs. The system will help applicants to assess themselves, and make choices concerning suitable careers and occupation. In addition, the system will assist in the upward mobility of current federal employees (OPM Operations Letter, 1989, pp. 2-3).

Summary and Recommendations

This article has shown the wide range of utilization of computers in public personnel assessment, from basic word processing applications to innovative item banking and computer administered work simulation tests. This area of personnel assessment is in its infancy in the public sector.

Recent technological advances in microcomputers (such as increased speed and storage capacity, and lower prices), the availability of reasonably priced microcomputer software for many applications such as item banking and adaptive testing, and advances in computer interactive video will likely lead to increased use of item banking and adaptive testing, and increased development and use of job-related and realistic personnel assessment devices which can be administered by computer. The medical work simulation examination briefly outlined in this article illustrates the future of computerized work simulation testing. However, use of these technologies involves costs which are beyond the currently available resources of many public sector jurisdictions.

The National Commission on Testing and Public Policy should recommend expanded support for the application of computer technology to assessment, and encourage interjurisdictional cooperation and innovation in this area.

USE OF TEST SCORES

Background Information

Many individuals feel that merit system procedures rely too much on tests and do not provide room for judgement in personnel selection. However, there is a trend in public sector assessment away from rigid use of test scores as the primary determiners of employment decisions to use of scores as aids in the decision-making process. This trend is evident in public sector practice and in guidelines and regulations.

A recent review of this trend is contained in the article "Should Personnel Selection Tests Be Used on a Pass/Fail, Grouping or Ranking Basis?" (Sproule, 1984). The article appeared in a special issue of *Public Personnel Management* on "Assessment Issues and Challenges." The paper describes legal requirements and professional standards relevant to the method of use of test scores, practices of some merit system agencies on their use of test scores; and provides guidance on how to decide on method of test score use, how to group test scores, and how to explain grouping of scores.

Summary of Requirements in Regulations, Laws, and Standards

"How test results are used can be as important a consideration in personnel selection as the type and content of tests used for hiring or promotion. This was recognized in the Uniform Guidelines on Employee Selection Procedures (UGOESP), (EEOC, et al., 1978). The section of the guidelines on 'Method of Use of Selection Procedures' states, in effect, that employers should use their selection procedures in accordance with the evidence of validity and utility supporting the selection procedure and should consider the degree of adverse impact of the method of use. Thus, the evidence may support pass/fail use of a test or other selection procedure, ranking may be appropriate, grouping or categorization of candidates may be appropriate, or some combination of these approaches may be appropriate.

(e.g., pass/fail on one test, and grouping on another)" (Sproule, 1984, p. 375).

"The American Psychological Association (APA) standards support use of the most valid methods available for decision-making and caution against use of subjective judgments which have not been validated. The standards caution that tests should not be interpreted too precisely and encourage use of confidence intervals in score reporting and consideration of important nontest information.

"The Division of Industrial and Organizational Psychology (APA) 'Principles' support use of tests on a ranking basis with some room for judgment in the use of scores.

"The federal merit system standards seem to encourage a wide latitude in methods of use and do not require use of scores on an absolute ranking basis. State and local laws and regulations have shown a trend from absolute ranking regulations (e.g., rules of one or three) to broader methods of use such as consideration of the top five or ten candidates or category systems." (Sproule, 1984, p. 379)

VIEWS ON THE MOST APPROPRIATE METHOD OF TEST SCORE USE

A variety of viewpoints are presented in the article on the most appropriate use of test scores. For example, articles by Millard, Wiesen, and the American Society for Personnel Administration are referenced which support strong reliance on test scores (Sproule, 1984, pp. 379-380).

On the other hand, some organizations, such as the NAACP Legal Defense Fund, see strict reliance on test scores as in the civil service ranking requirement of some public jurisdictions to be a "pretext for discrimination" (Goldstein, 1989).

"It is the author's opinion that the uniform guidelines provisions on method of use were intended to move employers away from ranking systems and heavy reliance on formal written test assessment approaches because of the adverse impact which paper and pencil tests typically have against minority groups. A difference of about one standard deviation in test scores is

typically found between minority and majority test performance. Because of these group differences in test performance, it is likely that there will be a greater degree of adverse impact against minorities when tests are used on a strict ranking basis, as compared to other methods of use, such as pass/fail utilization or broad grouping of scores. An employer will typically have more opportunity to consider affirmative action in hiring with pass/fail or broad grouping systems than with strict ranking.

"The best advice the author can provide to employers is to use test scores from well developed and job-related tests along with other relevant information in making selection decisions, and to generally make selections from among those with the highest scores when there is evidence which supports this. Use of test scores along with other data allows the employer to consider a variety of goals of the organization in making selection decisions, to consider job requirements not measured in the formal assessment process, and to match the candidates against the unique requirements of each job opening.

"The author believes that strict ranking should only be used when most of the important job requirements have been assessed and when there is very strong evidence of validity (e.g., criterion-related evidence or strong content evidence such as in a work-sample test). Grouping systems, which match the number of groups and the width of the score range of each group to the evidence supporting the test and the measurement method, would often be an appropriate method of use for merit system assessment devices. Pass/fail utilization of an entire selection procedure is, in the author's opinion, only applicable in very limited circumstances" (Sproule, 1984, p. 381).

PUBLIC SECTOR TRENDS

The article summarizes 1974, 1978, 1982, and 1983 surveys of merit system certification practices. A clear trend of movement from rules which allow consideration of only a

few top scoring candidates for a job opening (e.g., "rule of three") towards broader certification systems is evident in the survey results.

A few examples of broad certification systems follow:

"The District of Columbia uses grouping in most rating of training and experience examinations. Four groups of candidates are identified in the rating procedure: 'Ineligible,' 'Qualified,' 'Well Qualified,' and 'Highly Qualified.'

Illinois uses a category grade procedure where the top 25% of the candidates are assigned a grade of 'A,' the next 30% 'B,' the next 40% 'C,' and the bottom 5% an 'F' or fail rating. Michigan uses a rule of reliability or bandwidth certification where the range of scores certified depends on the reliability of the test" (Sproule, 1984, p. 383). Georgia allows appointments from the top ten whole scores.

Leo S. Goldstein of the New Jersey Department of Personnel reported on the use of "banding" instead of the "rule of three" for selection of candidates from a certification list with a large number of candidates. He stated: "Dividing the list into deciles will enable appointing authorities to select from a larger number (one-tenth the total) of certified candidates at one time. This may be helpful to minority candidates, who generally place lower on lists, by bringing their names up for earlier consideration" (Goldstein, Leo 1989).

Following is an example relating to the use of scores on a pass/fail basis. In 1988, the U.S. Office of Personnel Management announced a policy of use of college grade point average as one method of selection for entry level professional positions. Any candidate with a grade point average of 3.0 or above can be hired (U.S. Office of Personnel Management, 1988). This new policy is discussed in the rating of training and experience article in this paper.

There have been some legal decisions related to the use of test scores by merit systems. Some decisions have supported use of ranking, others have supported use of grouping, depending on the circumstances. For example, in 1980 the U.S. District Court (2nd cir.) ordered the New York City Civil Service Commission to stop using a Police Officer test

on a ranking basis because of a lack of test precision. In 1985, 1986, and 1987, the New York State Supreme Court issued decisions limiting the use of zone scoring by the New York State Department of Civil Service.

The report of the Volcker Commission appears to support the concept of flexibility in the use of test scores. For example, the report recommends: "...cabinet officers and agency heads should be given greater flexibility to administer their organizations, including greater freedom to hire and fire personnel" (National Commission on the Public Service, 1989, p. 20).

GUIDELINES FOR DECIDING ON HOW TO USE TEST SCORES

The article reviews considerations which can help in deciding on the most appropriate use of employment test scores. The following table outlines some of the factors described in the article which can be reviewed to help determine if a selection procedure should be used to rank candidates or group candidates into categories. The guidelines which are contained in the following table are not based upon empirical evidence. They are based upon principles of personnel assessment, provisions in the Uniform Guidelines on Employee Selection Procedures, and the judgment of the author. Readers interested in more details or an explanation of these guidelines should refer to the winter 1984 issue of *Public Personnel Management.*

The author does not generally recommend use of very broad categories in grouping or banding, such as considering all those who pass a test to be equal. The relationship between test scores and job success is generally linear (Griffin, 1989, pp. 128-129). Thus, some method of ordering or grouping applicants by score in making selection decisions is to the employers advantage.

Considerations in Deciding on Method of Use

Considerations	Reasons for Ranking	Reasons for Grouping
Job Structure	Work performed and requirements consistent from position to position.	Wide variance in work performed and requirements.
Candidate Pool	Many candidates, few jobs.	Few candidates, few jobs.
Adverse Impact	No history of adverse impact.	Past adverse impact.
Job Analysis	Complete, documented study conducted.	Limited job analysis information.
Proportion of Requirements Assessed	Large.	Few.
Testing Procedure	Multi-part test.	Single-part test.
Validity Evidence	Strong validity evidence in the literature for this test type. Content of test closely matches job behavior.	Procedure generally has low to moderate validity. Test content is general.
Test Precision	Test has large standard deviation and high reliability.	Test has small standard deviation and low or moderate reliability.
Remaining Selection Steps	No procedures to assess applicants further prior to hire.	Applicants will be assessed further prior to hire.

The journal article provides narrative guidance and examples on how to use test scores. For example, the article states:

> "There are few absolute rules on how to decide on the appropriate method of use. Professional judgment is needed in most instances. There is a variety of information which can help in making judgments on the proper method of use. As a general rule, the method of use should match the need for differentiating between candidates, the degree of coverage and precision of the assessment procedure, and the evidence on what is the appropriate type of differentiation" (Sproule, 1984, p. 384).

A recent IPMA "Handbook on Test Administration" also provides guidance on the use of test scores. Concerning the issue of ranking or banding, the handbook states:

> "It is the author's opinion that banding is a very useful method for both maximizing a test's utility and enhancing

test fairness. Straight rank-ordering would be preferable in situations with very strong evidence of test validity (empirical or otherwise), high reliability and a sufficient variance (spread) of scores" (Adams, 1989, p. 50).

SUMMARY AND RECOMMENDATIONS

There is a trend in public sector assessment away from rigid use of test scores as the primary determiners of employment to use of test scores as aids in the decision-making process. This trend is evident both in law and regulation and in public sector practice. However, many public sector jurisdictions still are bound by regulations which still require rigid reliance on test scores. Guidelines are available to help employers decide on the most appropriate method of use of test scores.

The National Commission on Testing and Public Policy should encourage the appropriate use of tests and test scores. Use of well developed tests should be encouraged. Tests can be helpful aids in the decision-making process and can contribute greatly to accurate prediction. Rigid reliance on test scores only in making employment decisions should be discouraged. Public sector agencies which have rigid laws or regulations on the use of scores, such as "rules of one" or "rules of three," should be encouraged to modify such provisions. The National Commission on Testing and Public Policy should recommend the development of public policy on the use of test scores.

LEGAL PROVISIONS TO ENCOURAGE INNOVATIONS AND RESEARCH AND DEMONSTRATION PROJECTS

BACKGROUND INFORMATION

In recent years, some legislative bodies have passed laws which allow and encourage waiver of merit system hiring requirements for research and demonstration purposes. Sometimes such projects are to experiment with alternative selection procedures, or to try out unique employment, recruitment, training or other techniques which may not be appropriate under existing regulations. Such flexibility can provide the opportunity to be innovative and creative. The National Commission on Testing and Public Policy should encourage legislative and administrative provisions which allow waiver of merit hiring requirements for research and demonstration purposes. Adequate funding to support such activities should be provided by governmental agencies.

MINNESOTA EXPERIMENTAL SELECTION PROJECTS

A very recent example is a provision for "Experimental Selection Procedures" in the state of Minnesota (Minnesota Department of Employee Relations, 1989). This 1987 state law provides that:

"...The Commissioner of employee relations may conduct experimental or research projects designed to improve recruitment, selection, referral, or appointment processes for the filling of state classified positions.... Any provision of Minnesota Statutes, ...associated personnel rules...or administrative procedures...is waived for the purpose of these projects. This waiver is limited to no more than five percent of appointments made.... The Commissioner shall report by March 1, 1988, and January 15, 1989, to the legislative commission on

employee relations the results of the experimental or research projects."

One experiment offered state hiring agencies the option to use job service referrals to fill entry-level typist and stenographer positions as an option to rank order eligible lists resulting from state civil service examinations. One hundred sixty of 179 hires studied were made by use of job service referrals; only 10.6% of hires were made from the regular eligible list. There was a slight (statistically insignificant) difference in "protected group" hires between job service referrals and eligible list appointments.

Hiring agencies were surveyed to obtain their opinions on the hiring option. They were more satisfied with the speed of the hiring process under the job service referral procedure and were also more satisfied with the quality of persons appointed from these referrals.

As a result of this experience, Minnesota is requesting legislation to incorporate this hiring option into statute. The state also plans to increase efforts to obtain greater referrals of protected group members from job service offices to agencies which have disparities in their work force.

Another experiment was conducted for seven classes of positions where promotion examinations are given. Under current law, Minnesota hiring agencies have the option of requesting a list of candidates limited to their own employees or a list of candidates from all agencies. Under the experiment, the number of candidates referred from all agencies for consideration was doubled from 10 to 20. This experiment was designed to provide more promotional opportunities to state employees and to encourage consideration of candidates outside the hiring agency.

This experiment has only recently begun and has not yet provided enough data for analysis. However, the concept has been positively received by hiring agencies and employee union representatives. The legislative committee has been requested to include authorization for this procedure in Minnesota statute.

Six experiments have been conducted or begun under this Minnesota law. A January 13, 1989, report to the Legislative Commission, "Employee Relations," concluded that:

"Overall, the experiment results have been successful. Some experiments have already shown considerable improvement over the standard selection or referral procedure. The opportunity to experiment itself has been judged the most successful. The option to conduct experiments has caused the staff to think more innovatively and imaginatively when confronting selection problems and opportunities.

"The concept of experimenting is an innovative idea worthy of continuation (and emulation by other states or local jurisdictions). This may be inferred from the selection of Minnesota's law as a semi-finalist in the Council of State Government's annual competition for 'Innovations.' The Department of Employee Relations will be requesting legislation to be permitted to continue to conduct experiments under similar controls and conditions on an ongoing basis."

Federal Civil Service Reform Act

A provision on research, demonstration, and other programs is contained in the Federal Civil Service Reform Act of 1978. This provision authorizes the U.S. Office of Personnel Management (OPM) to conduct and support personnel management research and to carry out up to ten demonstration projects at a time. Such projects can be by federal agencies, or by others funded by OPM under this provision, such as state or local governments, educational institutions or private organizations.

To date, one project has taken place in the employment selection and assessment area under the above provision. OPM's Office of Systems Innovation and Simplification is currently overseeing an applied project in the federal work force which authorized an alternative selection process for the employment of Air Traffic Controllers in the Federal Aviation Administration (FAA) of the Department of Transportation.

Previously, a written aptitude test was used to select Controllers. After the firing of Controllers by the President as a result of a strike, OPM authorized the waiver of regulations. "In 1982, FAA encouraged colleges and universities to

establish an Airways Science Curriculum, in an effort to up-date the educational levels of applicants seeking careers with FAA and to shift the costs of training individuals for civilian aviation occupations from government to the private sector."

Hiring under the new procedure is based upon the completion of an Airways Science Curriculum developed by colleges and universities under grant funds provided by FAA. This system is too new to have been fully evaluated. Many students in the program are just reaching the point where they will be considered for jobs. Only ten Airway Science graduates have been hired since the project started (Donna Beecher, personal communication and correspondence, February 23-24, 1989).

Another example of federal waiver of regulations was recently described in an OPM Operations Letter on the Office of Personnel Research and Development, as follows:

"OPM Regional Directors have recently been given authority to waive the written test portion of the clerical examination in shortage labor market situations. This authority was granted on the basis of preliminary results from a pilot test conducted recently by OPM's Boston Area Office. This pilot study revealed that waiving the test in shortage labor areas expedited the hiring process and that the quality of applicants hired under the waiver was acceptable" (U.S. OPM, 1989, p. 7).

Unfortunately, much of the federal funding for improvements in assessment procedures have been reduced drastically since the passage of this law. For example, the staff of OPM's Personnel Research and Development Center (now the Office of Personnel Research and Development) is about half of what it was in the late 1970s. Also, the staff of Personnel Research Psychologists in OPM's Regional Offices, who assisted federal agencies and state and local governments in efforts to improve assessment procedures, has been practically eliminated by funding cutbacks. Despite these cutbacks, OPM continues to produce a variety of innovations and products in the assessment area; but now provides practically no assistance to state and local governments in personnel assessment.

The National Commission on Testing and Public Policy should recommend increased federal, state, and local resources for assessment improvement efforts. Legal provisions for research and demonstration efforts need fiscal support to be meaningful.

Intergovernmental Personnel Act

Another example of a legal provision to encourage innovation is in the Intergovernmental Personnel Act (IPA) of 1970. The law authorizes federal grants to state and local governments to strengthen personnel administration, encourage innovation and allow for diversity. The law includes a provision for:

"...undertaking research and demonstration projects to develop and apply better personnel administration techniques, including both projects conducted by state and local government staff and projects conducted by colleges or universities or other appropriate nonprofit organizations under grant or contract..." and "strengthening the recruitment, selection, assignment, and development of handicapped persons, women, and members of disadvantaged groups whose capacities are not being utilized fully..." (United States Congress, 1971, Title II, Sec. 202 B. 5, p. 4)

Considerable progress in public sector assessment resulted from IPA projects. Examples of projects in state and local governments which were funded under this law are described in other parts of this paper. Some of these projects are: the work of Huett on improving the selection interview (described in the article on structured oral examinations); use of work sample tests in California for police, fire, and other occupations was supported and encouraged by an IPA project conducted from 1974 to 1976 (described by Joiner in the article in this paper on work sample tests); and the project by Farrell on work simulations in Minnesota (described in the article in this paper on work simulations).

In addition to the above, many other research and demonstration projects were conducted under the IPA. Most of the regional consortia of public personnel professionals

were formed with the support of IPA funding. A number of successful interjurisdictional test development and validation projects were conducted such as one for Conservation Officer in the New England states, and one for clerical occupations in the Mid-Atlantic area. Also, considerable assessment consulting help and training was provided to federal, state and local agencies by Psychologists working in the IPA program in regional offices of the U.S. Civil Service Commission.

Overall, the Intergovernmental Personnel Act contributed significantly to upgrading the skills and tools of assessment professionals in state and local governments. In the mid and late 1970s, this law and the technical assistance and research funding provided under it, led to many innovations and improvements in assessment in the federal government and in state and local governments.

Unfortunately, funds are no longer being provided for the above purposes under this Act. Mr. Ashton C. Morris, Acting Director, Intergovernmental Personnel Programs, U.S. Office of Personnel Management, informed the author in February 1989 that no funding or technical assistance has been provided under this Act for the last six or seven years. Mr. Morris is a one-person office which administers this Law.

The National Commission on Testing and Public Policy should encourage the federal government to fund and carry out the provisions of the Intergovernmental Personnel Act. The programs the law authorizes have fallen by the wayside in the budget cuts of the 1980s.

SUMMARY AND RECOMMENDATIONS

The federal government, and some state and local governments, have enacted legal provisions and administrative procedures to allow waiver of employment and training regulations for the purpose of research, demonstration, and innovation to improve selection and other personnel procedures. Examples of such regulations, and projects conducted under these laws, have been described in this article and other sections of this paper. Many such projects resulted in significant improvements in public personnel assessment. Unfortunately, federal funding support for innovations and research in public sector assessment has been drastically reduced in the federal service. Federal funding of programs to assist state and local governments under the Intergovernmental Personnel Act has been essentially eliminated, despite the past success of this program and the provisions of the law which provide for such assistance.

The National Commission on Testing and Public Policy should:

1. Encourage legislative and administrative provisions which allow waiver of merit hiring requirements for research and demonstration purposes.

2. Recommend increased federal, state, and local resources for innovation and assessment improvement efforts. Legal provisions for research and demonstration efforts need fiscal support to be meaningful.

3. Encourage the federal government to fund and carry out the provisions of the Intergovernmental Personnel Act.

EMPLOYMENT TESTING OF PERSONS WITH DISABILITIES AND EMPLOYMENT PROGRAMS FOR THE DISABLED

Background information

A June 1988 publication of the International Personnel Management Association Assessment Council (IPMAAC) illustrates the awareness, concern, and actions of public-sector assessment professionals in addressing the needs of particular groups in society by making appropriate modifications to job-related tests. The publication is a Personnel Assessment Monograph entitled: "Employment Testing of Persons With Disabling Conditions" (Daley, et al., 1988). This 60-page document goes considerably beyond the four pages of guidance on "Testing People Who Have Handicapping Conditions" in the "Standards for Educational and Psychological Testing" (American Educational Research Association, et al., pp. 77-80).

The IPMAAC publication begins with an article by Mary Ann Nestor of the United States Office of Personnel Management on the "Psychometric Implications of Test Modifications." This article describes test modifications by: 1) type of modification (e.g., time limit, content, test medium), and 2) type of disability (e.g., visual, motor, hearing, learning). Data is presented on the reliability and validity of modified educational tests. The reliability and validity data presented on modified tests is similar to that for unmodified tests. Dr. Nestor concludes with this statement on employment testing: "...careful accommodations that are made to maintain the measurement characteristics of the tests should lead to the retention of validity" (Daley, et al., 1988, p. 11).

The monograph then describes the New York State Department of Civil Service and the Pennsylvania State Civil Service Commission programs of providing testing accommodations. These case descriptions, by Michael Dollard of New York and Robert Schneider of Pennsylvania, il-

of New York and Robert Schneider of Pennsylvania, illustrate how employers can develop policies and carry out programs to modify various aspects of testing where appropriate. Information is provided on such topics as:

- how the programs were developed and are carried out,
- types and numbers of accommodations made,
- when accommodations are made,
- training of those involved in providing accommodations,
- the positive results and benefits of the programs, and
- problem areas.

Next, John Kraft of the U.S. Office of Personnel Management provides commentary on the New York and Pennsylvania programs, identifies sources of assistance and expertise in providing accommodations, comments on test accommodation issues, and identifies a situation where he recommends waiver of a test part.

Loren Daley of the Human Services Directorate, Canadian House of Commons, provides a discussion of test accommodations from a consumer's perspective. He comments on: the New York and Pennsylvania programs, problems associated with some accommodations, the self-esteem of disabled persons and the problems of self-identification; and advised to obtain the input of knowledgeable consultants and consumer/applicants. Daley sees tests as a barrier to employment and recommends "...the main effort should be devoted to finding jobs for disabled persons" (Daley, et al., 1988, p. 41).

Loren Daley's comment above clearly illustrates one of the issues which the National Commission of Testing and Public Policy faces. Tests are often viewed as the "barrier" to be overcome. The alternative proposed to tests is often to "find jobs" or "eliminate the test" for the particular group. The problem with these options is that decisions still need to be made in determining who is selected for the finite number of jobs. Whatever the decision-making strategy, the strategy is in effect a test to determine who is hired and who is not hired. Well developed and appropriately used tests

can contribute significant information to help in making such important decisions on a more objective, fair and job-related basis. As illustrated throughout this paper, there are assessment alternatives which do not have the high adverse impact of traditional paper and pencil tests.

One solution which assessment professionals offer in this dilemma is to make modifications in the use of tests so that tests are as appropriate and fair as possible for the particular group. In this instance, such modifications are designed to ensure that persons with disabilities are not disadvantaged when demonstrating their knowledges, skills and abilities on tests. The IPMAAC Monograph provides model guidelines on how to accomplish this. These guidelines were developed based on experience in test modification for persons with disabilities. A summary of the content areas of the guidelines follows.

MODEL GUIDELINES FOR ACCOMMODATED TESTING OF THE DISABLED

Appendix A of the IPMAAC Monograph contains model guidelines. These guidelines are seventeen pages in length and include:

- a proposed policy statement for organizations on accommodated testing. Following is a portion of the proposed policy:

"This agency will make reasonable accommodations to the known disabilities of candidates unless such accommodations would impose hardship on this agency's programs (Federal Rehabilitation Act of 1973 – Sections 503 and 504). Accommodating the individual disabled candidate's needs with the goal of maintaining the competitive nature of the selection process will be the primary concern. To this end, guidelines for the testing of those with various disabilities will be developed and revised from time to time by those specializing in testing and employment of the disabled. Exceptions to these general guidelines may be authorized by the (appropriate authority) to meet the needs of individual candidates.

"In general, the appropriate accommodation provided will be that which is normally used by or available to the candidate and can reasonably be considered as appropriate and available in the performance of the position being tested for" (Daley, et al., 1988, p. A-1).

— a definition of reasonable accommodations,
— guidance on obtaining and documenting information on a candidate's disability,
— guidance on the confidentiality of information on a candidate's disability,
— specific accommodation recommendations for a variety of disabling conditions, including a listing of test accommodation needs of persons in three categories of disabilities. The categorization briefly defines different types of disabilities and indicates the types of accommodation which are considered appropriate and reasonable for each.   The categories include: visually disabled (totally blind, partially sighted, and limited vision and special vision problems), hearing impaired, and disabled in processing information from the printed page.

Following is one brief example:

"Hearing impaired after development of normal language facility:

These candidates usually function in the same way as nondisabled candidates with respect to written material but must receive some accommodation with respect to oral test instructions. They should be routinely seated where they have a clear view of the monitors giving instructions.

"Monitors for hearing impaired candidates should be screened and briefed on their responsibilities.

"Reasonable accommodations:

— written instructions for all parts of the testing sessions
— interpreter

— special seating as appropriate" (Daley, et al., 1988, p. A-7).

Guidelines are also provided on a variety of other topics such as: making test modifications, administering and scoring modified tests, accommodating nonwritten tests, testing time allowances and testing sessions needed, responsibilities and sharing of information (e.g., obtaining the assistance of the Affirmative Action Officer), providing alternate test dates, and accommodations for temporary disabilities.

In summary, the IPMAAC Monograph illustrates the efforts of the assessment profession to address the needs of particular groups in society and to take action to ensure the appropriate and fair use of employment tests. The National Commission on Testing and Public Policy should recommend the adoption of model guidelines for accommodated testing of the disabled by employers and educational institutions.

Another option is adoption of social policy goals which attempt to address both the needs of particular groups for employment, and the goal of employment on the basis of relative merit. The Georgia and Canadian programs which follow are examples of efforts to address both policy directions.

ALTERNATIVE EMPLOYMENT PROGRAM

In October 1988, the state of Georgia established a pilot "Alternative Employment Program" for applicants with severe physical impairments. Georgia's program incorporates both the approach of providing jobs advocated by Daley in the monograph described earlier, and the approach of providing testing accommodations described by Nestor, Dollard, Schneider, and Kraft. A document describing the personnel policy issues and actions leading up to this program is available from the state of Georgia.

"The process uses Rehabilitation Employment Specialists – who are trained in the assessment of disabled applicants and in job analysis – for position specific 'matching' with appropriate accommodation. After appointment there is a twelve-month working test period for on-the-job assessment of performance. This

process is in lieu of the centralized competitive program of accommodated test administration. However, both the competitive and the alternative procedures are available as processing 'options' for applicants with physical disabilities" (Maye, 1988).

OTHER PUBLIC SECTOR REFERENCES AND PROGRAMS

U.S. Federal Publications

The Bureau of Intergovernmental Personnel Programs of the U.S. Civil Service Commission (now the Office of Personnel Management) issued two useful publications to assist employers. One publication is designed to increase understanding of disabilities. Another provides guidance on setting up programs for improving employment opportunities for the disabled.

The publication "Employment of the Handicapped in State and Local Governments: A Guide to Specific Disabilities" briefly describes various disabilities and gives data on the number of handicapped persons in these categories: blind and visually handicapped, deaf and hard-of-hearing, other physical handicaps, the mentally retarded, and the mentally restored. For each category, the publication gives guidance on job possibilities, testing accommodations, job orientation or modification, and other topics (U.S. Civil Service Commission, 1983).

Canadian Federal Publications and Programs

The Public Service Commission of Canada has developed a 26 page "Personnel Management Manual Chapter" on "Disabled Candidates." The manual states:

"When necessary, reasonable accommodations must be provided to applicants who have physical, mental, psychiatric, or learning disabilities, to ensure that they are afforded opportunities equal to those of other applicants to demonstrate their qualifications for positions" (Public Services Commission, 1988).

144

The Canadian program includes an "Employment Equity Coordinator" in each of the Public Service Commission's offices. Each Coordinator is responsible for: recruitment of qualified disabled persons, seeking employment opportunities in Canadian federal agencies on behalf of applicants with disabilities, recommending accommodations, matching candidates to job openings, referring candidates for interviews and advising and assisting departmental managers.

An "Access Program for Disabled Persons" is conducted to provide on-the-job training in federal agencies for disabled persons who lack sufficient work experience. The costs of the trainee's salary can be reimbursed to the department under this program. Some Canadian federal agencies have set aside funds to purchase technical aids needed by disabled persons hired.

The Personnel Psychology Center of The Public Service Commission is responsible for making or determining appropriate test modifications for disabled persons. The Personnel Management Manual contains a section on "General Principles and Procedures for Assessing Disabled Persons." In addition, guidelines for testing persons with disabilities are contained within a section of the Public Service Commission's manual on "Testing in the Public Service of Canada."

A variety of Canadian federal agencies participate in the program. For example, the Translation Bureau of the Secretary of State provides sign language interpreters for deaf applicants and employees.

The Public Service Commission program includes a variety of features to protect the confidentiality of information on candidates and their disabilities.

Within the Canadian federal service, "Exclusion Approval Orders" allow the government to set aside the merit requirements of the law, with the approval of the Governor in Council, in order to address the issue of placing designated group members in particular jobs. The Public Service Commission has issued a set of guidelines concerning this process. The guidelines define the roles and responsibilities of federal departments, unions and the Public Service Commission in this process. For example:

"Departments must demonstrate with quantifiable goals that it does not serve the best interests of the Public Service, i.e., for reasons of efficiency, sensitivity and responsiveness to the public, equality of access to Public Service employment, equity in the treatment of all employees, or other valid objectives, to apply part or the whole of the Act..." (Public Service Commission, Guidelines, p. 4).

SUMMARY AND RECOMMENDATIONS

This article illustrates the efforts of the public sector assessment profession to address the needs of particular groups, as well as actions taken to help ensure the appropriate and fair use of employment tests. In addition, examples of programs to provide jobs for the disabled and other groups were identified.

The National Commission on Testing and Public Policy should recommend the adoption of guidelines on test modification and test accommodation by employers and educational institutions. The Commission should also encourage public policy makers to more directly address the needs of particular groups.

REFERENCES

Abrams, Nancy E. (1985). Alternative measurement methods and methods of use. *Instructor's Manual: Examination Planning Seminar*. International Personnel Management Association Assessment Council, Washington, D.C.

Abrams, Nancy E., Ash, R. A. and Johnson, J. C. (1986). *Instructor Manual and Participant Manual: Seminar on Ratings of Training and Experience*. International Personnel Management Association Assessment Council, Alexandria, Va.

Abrams, Nancy E., Laguardia, L. M. and Sheibley, L., (1981). How to develop job-related minimum qualifications and ratings of training and experience. Workshop presented at the Annual Conference of the International Personnel Management Association Assessment Council, Denver, Colo.

Adams, Joanne (1989). *Handbook on Test Administration*. International Personnel Management Association, Alexandria, VA.

Amick, Cynthia (1980). Summary of selected court decisions related to minimum qualifications. Paper from seminar on minimum qualifications and ratings of training and experience, Mid-Atlantic Personnel Assessment Consortium.

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, (1985). *Standards for Educational and Psychological Testing*. American Psychological Association, Washington, D.C.

Ash, Ronald A. (1980). The activity/achievement indicator: A possible alternative to the behavioral consistency method of training and experience evaluation. *Public Personnel Management*. Vol. 15, No. 3, 325-343.

Baker, Dale (1989). Assessment centers: Alive and well. Operations Letter. Office of Personnel Research and

Development, U.S. Office of Personnel Management, Washington, D.C., No. 337-1503, 4-5.

Barrett, Gerald V., et al. (1978). Values and professional judgement in validating and litigating tests for civil service positions. *Professional Psychology.* American Psychological Association, Washington, D.C.

Berkley, Steve (1984). *Appendix 67, Validation Report: Corrections Officer Trainee.* Pennsylvania State Civil Service Commission, Bureau of Personnel Assessment.

Bretz, Robert D. (1989). College grade point average as a predictor of adult success: A meta-analytic review and some additional evidence. *Public Personnel Management,* Vol. 18, No. 1, 11-22.

Bullock, Carol (1989). Increasing the odds for producing valid biodata instruments. *Proceedings of the 1983 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 76-77.

Caplan, J. and Schmidt, F. (1977). The validity of education and experience ratings. Paper presented at Annual Conference of the International Personnel Management Association Assessment Council, Kansas City, Mo.

Campion, Michael A., Pursell, E. D., and Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology,* Vol. 41, No. 1, 25-42.

Cline, Catherine (July 1988). Multiple-choice in-baskets for managers. *MAPAC News,* 4-5.

Colberg, Magda (1985). Logic-based measurement of verbal reasoning: A key to increased validity and economy. *Personnel Psychology,* Vol. 38, No. 2, 347-359.

Colberg, Magda, and Nester, Mary Ann (1987). The use of illogical biases in psychometrics. U.S. Office of Personnel Management. Paper presented at the convention of the American Psychological Association, New York City, N.Y.

Commonwealth of Pennsylvania, (1982). Development and Validation of Classification Standards. Governor's

Office, Secretary of Budget and Administration, Management Directive 520.7

Cook, Cindy Lorentson (1980). Rating education, training, and experience in the public sector. *Proceedings of the 1980 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 16.

Corcione, Glenda K. (1985). *Alternative Selection Methods: Biodata – A Viable Option.* New York State Department of Civil Service, Albany, N.Y.

Corcione, Glenda K. (1985). Biodata as an alternative selection technique: An extensive evaluation. *Proceedings of the 1985 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 101-106.

Corcione, Glenda K., and Means, Robert (1986). Biodata research project: the new york state experience. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 49-52.

Daley, Lorne, Dollard, M., Kraft, J. Nestor, M. and Schneider, R. (1988). Employment testing of persons with disabling conditions. International Personnel Management Association Assessment Council, *Personnel Assessment Monograph,* 1 (4).

Darany, Theodore S. (1980). Automating the examination process in the 80s. Presidential Address. *Proceedings of the 1980 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 35.

Darany, Theodore S. (1984). Computer applications to personnel (Releasing the genie–Harnessing the dragon). *Public Personnel Management,* Vol. 13, No. 4, 451-473.

Darany, Theodore S. (1986). Computer assisted proctoring: a better way to administer tests. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 135-138.

Darany, Theodore S. and French, Jennifer (1986). *Development of the Western Region Item Bank.* Department of Personnel, San Bernardino County, Ca.

Davey, Bruce W. (1986). A computer-administered interest inventory. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 143-148.

Davey, Bruce W. (1986). Personnel testing and the search for alternatives. *Public Personnel Management,* Vol. 13, No. 4, 361-374.

Davey, Bruce W. and Duffy Wallace, Karen (1986). Development of a high-structure, competency based oral exam for police sergeants. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 148-152.

Davis, Roger (1986). Raising the validity of the oral examination: The BOSS technique. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 153-157.

Delaware State Personnel Office (1988). *Knowledge, Skill and Ability Based Minimum Qualifications: A Guide for Developers.*

Dye, David A. and Reck, Martin (1988). *A Literature Review and Meta-Analysis of Education as a Predictor of Job Performance.* Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, D.C.

Dye, David A. and Reck, M. (1989). College grade point average as a predictor of adult success: A reply. *Public Personnel Management,* Vol. 18, No. 2, 235-241.

Dye, David A., Reck, M. and McDaniel, M. A. (1989). Moderators of the validity of written job knowledge measures. U.S. Office of Personnel Management. paper presented at the Fourth Annual Conference of the Society of Industrial and Organizational Psychology.

Eisenhart, R. J. (1986). *Computers in Testing; Survey Results: Computer Use / Assistance in Selection and*

Examinations Operation. Personnel Assessment Research Division, Pennsylvania State Civil Service Commission, Harrisburg, Pa.

Ellison, Robert L. (1971). An alternative to discrimination in testing? biographical data as a selection device. Presentation to the National Civil Service League, Institute for Behavioral Research in Creativity, Salt Lake City, Utah.

Fales, Roy (1989). *Trades Performance Examinations and Testing Center.* New Jersey Department of Personnel.

Farrell, Barry A. (1979). *Job-Specific Assessment Simulation.* Selection Research Unit, Minnesota Department of Personnel.

Faust, Wayne (1988). News from the southern region—work sample testing. *Assessment Council News.* International Personnel Management Association Assessment Council, Vol. 12, No. 4, 5-6.

Ferrara, Philip (1981). The structured interview—dead or alive? *Proceedings of the 1981 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 43.

Flynn, John T., Anderson, B. and Rubovits, J. J. (1984). The development of reliable oral interview procedures for promotional candidates. *Proceedings of the 1984 International Personnel Management Association Assessment Council Conference,* 26-28.

Friedlander, G. H. (1988). Civil service goes to the movies. Paper presented at the 1988 Annual Conference of the International Personnel Management Association Assessment Council, Las Vegas, Nev.

Goldstein, Barry (1989). A plaintiff's attorney's view of recent judicial decisions regarding employee selection and affirmative action. Oral presentation to the Mid-Atlantic Personnel Assessment Consortium Winter Meeting, Washington, D.C.

Goldstein, Leo S. (1988). Atlantic City expressway project. Unpublished report. New Jersey Department of Personnel.

Goodison, Marlene (1980). *A Summary of 10 Cities On-Site Interviews and a Critique of CODAP (Comprehensive Occupational Data Analysis Program).* Center for Occupational and Professional Assessment, Educational Testing Service, Princeton, N.J.

Gordon, Stephen M. and Siracusa, Joseph (January 1989). New exam format for fire officials. *MAPAC News,* New Jersey Department of Personnel, 6-7.

Griffin, Margaret E. (1989). Personnel research on testing, selection, and performance appraisal. *Public Personnel Management,* Vol. 18, No. 2, 127-137.

Guion, Robert M. and Gibson, Wade M (1988). Personnel selection and placement. *Annual Review of Psychology,* Vol. 39, 349-374.

Harris, Patricia A (1988). *The Effectiveness of Grade Point Average and Level of Education in Predicting Performance in the Work Place: A Lay Summary of Findings.* Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, D.C., OPDR Report 88-1.

Hogan, James B. and Stokes, Garnett S. (1989). The Influence of socially desirable responding on biographical data of applicant versus incumbent samples: Implications for predictive and concurrent research designs. Paper presented at the Fourth Annual Conference of the Society for Industrial and Organizational Psychology, Boston, Mass.

Horney, Nicholas F. and Dertien, Marvin G. (1981). Job evaluation through computer use. *Proceedings of the 1981 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 19-20.

Huett, Dennis L. (1976). Improving the selection interview in a civil service setting. Paper presented at Great Lakes Assessment Council and International Personnel Management Association, Chicago, Ill.

Hunter, John E., and Hunter, Ronda F. (1983). *The Validity and Utility of Alternative Predictors of Job Performance.*

Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, D.C., OPRD-83-4.

International Personnel Management Association Assessment Council (1981). *Bylaws.* Washington, D.C.

Jacobson, Larry S. (1986). Computerized simulation testing: A basic language program to develop and automate simulation tests. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 138-143.

Johnson, James C., Guffey, W. L. and Perry, R. A. (1980). When is a T and E rating valid? Paper presented at the Annual Meeting of the International Personnel Management Association Assessment Council, Boston, Mass.

Johnston, S. (1982). *Guide for Class Specification Review.* Bureau of Personnel Assessment, Pennsylvania State Civil Service Commission.

Joiner, Dennis A. (1984). Assessment centers in the public sector: A practical approach. *Public Personnel Management,* Vol. 13, No. 4, 435-450.

Joiner, Dennis A. (1987). Content valid testing for supervisory and management jobs: A practical/common sense approach. *Personnel Assessment Monographs.* International Personnel Management Association Assessment Council, Vol. 1, No. 3.

Joiner, Dennis A. (1987). Orientation to assessment centers: A video approach. *Proceedings of the 1985 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* pp. 172-173.

Kaido, Betsy (1988). East coast dispatch – police testing. *Assessment Council News.* International Personnel Management Association Assessment Council, Vol. 12, No. 4, 7.

Kaido, Betsy (1988). Test Automation. *Assessment Council News.* International Personnel Management Association Assessment Council, Vol. 12, No. 3, 5.

Kaiser, Paul (1988). Video testing: What you see is what you get. Paper presented at the 1988 Annual Conference of

the International Personnel Management Association Assessment Council, New York State Department of Civil Service.

Kehew, David L. (1982). Evaluation standards and procedures. *Examiner's Manual.* Bureau of Personnel Assessment, Pennsylvania State Civil Service Commission, Harrisburg, Pa., Chap. 4.

Kleiman, Lawrence S., Falls, V. C. and Wilcox, D. L. (1982). Personal life history—Biodata items as suitable alternative selection procedures: A problem and proposed solution. *Proceedings of the 1982 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 58-59.

Kraus, John C. (1986). The multiple-choice in-basket exercise as developed and used by the New Jersey Department of Civil Service. *Proceedings of the 1986 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 46-49.

Lacomba, Lourdes (1983). Training and experience evaluations and other forms of self-assessment: Assessment for the selection and recruitment of public health nurses. *Proceedings of the 1982 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 58-59.

Lawton, Jack R. (1982). *A Study of Return Rates and Applicant Qualifications in the Behavioral Consistency Model.* Wisconsin Department of Industry, Labor and Human Relations.

Lawton, Jack R., Showers, Barbara A. (1980). Writing assessment for merit systems employment. Wisconsin Department of Employee Relations. Paper presented at the Annual Conference of the International Personnel Management Association Assessment Council.

Linsley, Clyde (September 1988). Playing the recruitment game. *Government Executive,* 42-45.

Long, Ernie (1984). Computerized ratings of training and experience. *Proceedings of the 1984 International Person-*

nel Management Association Assessment Council Conference on Public Personnel Assessment, 17-20.

MacLane, Charles N. (1979). *A Work-Sample Approach To Selection for Upward Mobility Positions.* Technical Study 79-2, Personnel Research and Development Center, U.S. Office of Personnel Management, Washington, D.C.

MacLane, Charles N. and O'Leary, Brian S. (1988). Job specific tests and an overview of research on alternatives. Paper presented at the Annual Meeting of the American Psychological Association in Atlanta. Prepared by staff of the Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, D.C.

Mann, Walter G. (1983). Feasibility of a content-valid biographical questionnaire for the selection of municipal Police Officers. *Proceedings of the 1983 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 77-78.

Mann, Walter G. (1988). *Validation of Apprentice Examinations: Criterion-Related Validity.* Office of Personnel Research and Development, U.S. Office of Personnel Management, Washington, D.C.

Maye, Doris M. (1988). Selection procedures for applicants with physical disabilities. *Personnel Assessment Sources, PASS* – A newsletter published by The International Personnel Management Association Assessment Council, Vol. 13, 2.

Mayes, Phyllis (1980). Computer-based job analysis: Some innovative applications to personnel management. *Proceedings of the 1980 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 14.

McClung, Glenn G. (1985). Problems and payoffs in automated applicant tracking. *Proceedings of the 1985 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 29-32.

McDaniel, Michael A. (1987). The staffing decision support system: Computer–assisted support for staffing decision makers. Paper presented at the Annual Conference of the International Personnel Management Association Assessment Council.

McDaniel, Michael A., and Schmidt, Frank L. (1985). *A Meta-Analysis of the Validity of Training and Experience Ratings in Personnel Selection,* Office of Staffing Policy, U.S. Office of Personnel Management, Washington, D.C., OSP-85-1.

McDaniel, Michael A., and Schmidt, Frank L. (1989). Computer-assisted staffing systems: The use of computers in implementing meta-analysis and utility research in personnel selection. *Public Personnel Management,* Vol. 18, No. 1, 75-86.

McDaniel, Michael A., Schmidt, F. L. and Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology,* Vol. 41, No. 2, 283-314.

McGuire, Janet (1982). Critical incident oral examining: A technique for improving fire and police promotionals. *Proceedings of the 1982 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 44.

McGuire, Janet (1986). Clerical work samples: Three practical approaches to scoring. Paper presented at the International Personnel Management Association Assessment Council Annual Conference, Psychological Services, Inc., Washington, D.C.

McKillip, Richard H. and Clark, Cynthia L. (1974). *Biographical Data and Job Performance.* Personnel Research and Development Center, U.S. Civil Service Commission, Washington, D.C.

Minnesota Department of Employee Relations (1989). *Report on Experimental Selection Projects Pursuant to Laws of 1987.* Submitted to Legislative Committee on Employee Relations, Chap. 186, Sec. 14.

Moreano, A. G. and Sproule, C. F. (1976). *Reliability and Other Data on Structured Oral Examinations.* Pennsylvania State Civil Service Commission, Bureau of Examinations, Harrisburg, Pa.

Myers, David C., et al. (1983). Computer-assisted testing in government and industry. *Proceedings of the 1983 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment, 75.*

National Commission on the Public Service (1989). *Leadership for America: Rebuilding the Public Service,* Washington, D.C.

New York State Department of Civil Service (1987). *Eligible List Evaluation Project: Motor Vehicle Representative III Computer-Assisted Examination.* Albany, N.Y.

New York State Department of Civil Service (1988). *Evaluating the Use of Biographical Data for Screening Therapy Aide Candidates: A Second Look.* Policy Analysis and Research Division, Albany, N.Y.

New York State Department of Civil Service (undated). *Written Simulations and Their Uses.* Paper describing latent image simulation testing, Albany, N.Y.

Nettles, Steve (1981). The structured interview—Dead or alive? *Proceedings of the 1981 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment, 41-43.*

Ocasio, Ben (1988). *An assessment center for postal inspector.* Presentation by the U.S. Postal Service to the Mid-Atlantic Personnel Assessment Council, Washington, D.C.

Plumlee, Lynnette B. (1980). *A Short Guide to the Development of Work Sample and Performance Tests* (2nd ed.). Personnel Research and Development Center, U.S. Office of Personnel Management, Washington, D.C.

Public Service Commission of Canada. *Guidelines Concerning the Approval Process of Exclusion Approval Orders by the Governor in Council.* Memorandum to the Commission, file no. 654-3-79-3-1.

Public Service Commission of Canada (1988). *Personnel Management Manual: Disabled Candidates,* Vol. 6, Chap. 12-8, Amendment/Modification 88-1.

Reilly, Richard R. and Wrech, Michael A. (1988). Draft paper: The validity and fairness of alternatives to cognitive tests. Prepared for the National Commission on Testing and Public Policy.

Richardson, Bellows, Henry & Co., Inc. (1988). *Technical Report: The Law Enforcement Officer Candidate Record.* Washington, D.C.

Schmidt, Frank L. and Hunter, John E. (1981). Employment testing: Old theories and new research findings. *American Psychologist,* Vol. 36, No. 10, 1128-1137.

Schultz, Chuck, Crowley, D., Anderson, G. and Patin, D. J. (1983). *Oral Examination Development Manual* (2nd ed.). Washington State Department of Personnel.

Sheibley, Helen Chattin (1986). Oral/work simulation testing: An alternative to standard assessment center testing. Paper presented at the 1986 Annual Conference of the International Personnel Management Association Assessment Council.

Sheibley, Leroy (1980). *Minimum Qualifications.* A seminar on minimum qualifications, and ratings of experience and training. Mid-Atlantic Personnel Assessment Consortium.

Silverman, Stanley B. and Wexley, Kenneth N. (1987). The selection interview: Some reasons for optimism. *Personnel Assessment Monograph.* International Personnel Management Association Assessment Council, Vol. 1, No. 3.

Sproule, Charles F. (1988). Initial use of personal computers in a personnel assessment program. *Assessment Council News,* Vol. 12, No. 3, 10-12.

Sproule, Charles F. (1984). Should personnel selection tests be used on a pass-fail, grouping or ranking basis? *Public Personnel Management,* Vol. 13, No. 4, 375-394.

Templer, Andrew J. and Tolliver, James M. (1983). Affirmative action—Making it effective in the public sector. *Public Personnel Management,* Vol. 12, No. 2, 211-217.

Trattner, Marvin H., Lockwood, R. E., Mayes, P. and Frank, M. (1980). Computer-based job analysis —Some innovative applications to personnel management. *Proceedings of the 1980 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment,* 13-14.

Tomes, William (1987). *Realistic job preview—Corrections Officer.* Presentation at the Winter Meeting of the Mid-Atlantic Personnel Assessment Consortium.

U.S. Civil Service Commission (1983). *Employment of the Handicapped in State and Local Government: A Guide to Specific Disabilities.* Bureau of Intergovernmental Personnel Programs, BIPP 152-83.

U.S. Civil Service Commission (1977). *Examination Preparation Manual.* Bureau of Recruitment and Examining, Washington, D.C.

U.S. Merit System Protection Board (1987). *In Search of Merit: Hiring Entry-Level Federal Employees.* Report to the President and the Congress of the United States, Washington, D.C.

U.S. Merit System Protection Board (1988). *Attracting Quality Graduates to the Federal Government: A View of College Recruitment.* Report to the President and the Congress of the United States, Washington, D.C.

U.S. Office of Personnel Management (March 6, 1980). *Alternative Selection Procedures.* Federal Personnel Manual, FPM Bulletin No. 331-3, Washington, D.C.

U.S. Office of Personnel Management (June 22, 1988). *New Program to Fill GS-5 and 7 Entry-Level Jobs.* Office of the Director.

U.S. Office of Personnel Management (January 27, 1989). *Office of Personnel Research and Development (OPRD) Updates.* Operations Letter 337-1503.

United States Congress (1978). *Civil Service Reform Act.*

United States Congress (1971). *Intergovernmental Personnel Act of 1970*. Public Law 91-648, 91st Congress, S. 11.

Valadez, Christina L. (1987). *Improving the interview*. Washington State Department of Personnel, Test Development.

van Rijn, Paul (1980). *Biographical Questionnaires and Scored Application Blanks in Personnel Selection*. Personnel Research Report 80-31, Personnel Research and Development Center, U.S. Office of Personnel Management, Washington, D.C.

Wiesner, Willi H. and Cronshaw, Steven F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, Vol. 61, 275-290.

Whetzel, Deborah L., McDaniel, M. A. and Schmidt, F. L. (1985). The validity of employment interviews: A review and meta-analysis. Office of Examination Development, U.S. Office of Personnel Management, 87-9, Washington, D.C..

White, Karen O. (1982). Update on the use of the behavioral consistency approach to unassembled examining in the federal government. *Proceedings of the 1982 International Personnel Management Association Assessment Council Conference on Public Personnel Assessment*, 18-19.

Wright, Patrick M., Lichtenfels, P. A. and Pursell, E. D. (1988). *The Structured Interview: Additional Studies and a Meta-Analysis*. University of Notre Dame, Center for Research in Business, Research Paper Series 88-07.

ipmaac

April 1990