

---

---

# *Personnel Assessment Monographs*



Employment Testing:  
A Public Sector Viewpoint



*Volume 2, Number 3  
December 1990*

*ipmaac*

---

---

# **E**mployment Testing: A Public Sector Viewpoint

Joel P. Wiesen, Ph.D.  
Massachusetts Department of Personnel Administration and  
Applied Personnel Research  
Newton, Massachusetts\*

Nancy Abrams, Ph.D.  
Personnel Consultant  
Fairport, New York\*

Sally A. McAttee, Ph.D.  
Milwaukee Department of Employee Relations  
Milwaukee, Wisconsin\*

Personnel Assessment Monographs is published by the Assessment Council of the International Personnel Management Association (IPMAAC), 1617 Duke Street, Alexandria, Virginia 22314. A copy is provided without charge to members of the Assessment Council. Others may obtain copies by writing the Director of Assessment Services at the above address. The monographs are intended to provide the most current information available on topics of broad interest in the field of applied personnel assessment. Manuscripts and proposals are solicited by the editor. Most appropriate are reviews of research on assessment methods, and innovative applications and solutions to personnel assessment problems in employee selection, performance appraisal, program evaluation, management analysis and related areas which are not generally available elsewhere. Manuscripts are reviewed by the editor and by consulting editors selected for their expertise in the area covered by a particular paper. They are selected for publication based on the technical and professional merit of the work, its applicability to current issues, and its utility and interest to the targeted readers. Inquiries, proposals and manuscripts may be sent to the editor, Jeffrey P. Feuquay, Oklahoma Office of Personnel Management, Jim Thorpe Building, 2101 North Lincoln Boulevard, Oklahoma City, Oklahoma 73105.

Jeffrey P. Feuquay, Ph.D., Editor

---

\* At the time the paper was written, in order of listing, the authors were the current President, the Past-President and the President-Elect of the International Personnel Management Association Assessment Council.

---

## *Editor's Preface*

---

Employment Testing: A Public Sector Viewpoint began as an effort of officers of the International Personnel Management Association Assessment Council to communicate to the National Commission on Testing and Public Policy (NCOTAPP) the uniqueness of public sector employment testing. The Editor thanks NCOTAPP for accepting this input, and for the opportunity to publish the paper as a monograph. Given the independent editing of the original paper, readers should expect some differences between the NCOTAPP and Monograph versions.

The Editor thanks Doctors Wiesen, Abrams and McAttee for facilitating the editorial process through their continuing suggestions and revisions; the synergism was a productive pleasure. Also due credit are the many staff members of the Oklahoma Office of Personnel Management who supported bringing the paper to press.

### **Joel P. Wiesen, Ph.D.**

Joel P. Wiesen is an industrial psychologist. He has spent most of his professional career researching and developing valid and practical personnel assessment and selection programs and procedures in both the public and private sectors. He has directed the civil service test development and validation program for all municipal and state civil service positions in the Commonwealth of Massachusetts. As an expert witness, he has been involved in numerous employment discrimination court cases and administrative proceedings. After working full-time for the Commonwealth of Massachusetts for 10 years he took advantage of the Commonwealth's alternative work option program. He now works part-time for the Commonwealth where he directs the research and evaluation efforts of the Massachusetts Department of Personnel Administration, lectures at Northeastern University, and heads his own consulting firm, Applied Personnel Research, located in Newton, Massachusetts.

### **Nancy E. Abrams**

Nancy E. Abrams is a former President of IPMAAC and is now a member of the Board of Directors, serving as IPMAAC's representative on the IPMA Executive Council. Nancy received her Ph.D. from Columbia University in Psychology: Measurement and Evaluation. She was the Principal Consulting Psychologist for the New York region of the U.S. Office of Personnel Management for a number of years. Since 1981, Dr. Abrams has been a self-employed consultant specializing in personnel management and measurement, working with a wide variety of public and private sector organizations.

### **Sally A. McAttee**

Sally A. McAttee is Staffing Services Manager for the city of Milwaukee, where she is responsible for personnel recruitment and selection. She has directed the development of and conducted research on experience ratings, including behavioral consistency ratings. She was previously the head of test validation for the state of Michigan, where she began her assessment career in 1967. She has a Ph.D. in Measurement and Evaluation from Michigan State University and is past president of the International Personnel Management Association Assessment Council.

---

## *Acknowledgments*

---

The authors are honored and pleased that the National Commission on Testing and Public Policy accepted an offer made by the Assessment Council, a professional section of the International Personnel Management Association, to provide this paper detailing aspects of assessment which are unique to employee selection in the public sector. At the same time, the views here are those of the authors and do not reflect the positions of the International Personnel Management Association Assessment Council (IPMAAC), the International Personnel Management Association, or the government agencies where two of the authors are employed.

Joel P. Wiesen, Ph.D.  
Nancy Abrams, Ph.D.  
Sally A. McAttee, Ph.D.

May 1989

# Personnel Assessment Monograph

Employment Testing: A Public Sector Viewpoint

Volume 2, Number 3

December 1990

## CONTENTS

<b>Introduction</b> . . . . .	1
<b>The Public Sector Context</b> . . . . .	2
Merit Principles . . . . .	2
Common Characteristics of Federal, State and Local Governments . . . . .	8
Differences Among Federal, State and Local Civil Service Systems . . . . .	11
Comparison of Selection in the Public and Private Sectors . . . . .	16
Some Factors Shaping and Constraining Personnel Selection In the Public and Private Sectors . . . . .	18
<b>Special Conflicts In Public Sector Testing</b> . . . . .	20
Conflicting Statutory Mandates . . . . .	20
Conflicting Statutory and Political Mandates . . . . .	24
Conflicting Legal and Professional Standards . . . . .	25
Pressures from Applicant Groups . . . . .	30
Conflicting Loyalties of Human Resource Professionals . . . . .	31
Factors Contributing to Conflict . . . . .	32
Conclusion . . . . .	35
<b>Constraints On The Use of Testing Methods</b>	
<b>In The Public Sector</b> . . . . .	36
Test Techniques . . . . .	37
Similarities to the Private Sector . . . . .	42
Types of Tests Rarely Used in the Public Sector . . . . .	43
<b>Areas of Greatest Need for Psychometric Research</b> . . . . .	45
Measurement of the Affective Domain . . . . .	45
Selecting Supervisory and Management Personnel . . . . .	47
<b>Call for Systematic Funding of Research and Development</b> . . . . .	57
Need for Legislation Encouraging Research . . . . .	57
Need for Research Funding . . . . .	58
<b>Conclusion</b> . . . . .	60
Summary . . . . .	60
Recommendations . . . . .	61
<b>References</b> . . . . .	63

---

## *Introduction*

---

Unique legal mandates and public expectations have led the public sector to develop personnel assessment, selection and promotion methods and systems which are unknown in the private sector or differ substantially from private sector practices. This paper describes some of the reasons assessment in the public sector is unique, explains some of the related practical constraints and psychometric implications, evaluates testing methods in light of these, and identifies some of the most pressing unmet assessment needs of the public sector. The monograph presents a practitioner's view of the assessment issues before the public sector: (a) the context in which testing operates in the public sector, (b) special conflicts in public sector personnel testing, (c) the promise, status, and technological, practical and political constraints of various current selection methods, and (d) some areas which most cry out for new and improved assessment methods. Finally, suggestions are made for improving public sector personnel assessment and selection.

This paper will focus on the type of personnel assessment which typically is the basis for hiring of civil service employees in the public sector, excluding the military. A similar system is often used for promotion.

---

## ***The Public Sector Context***

---

Testing in the public sector is shaped, directed and constrained by legal, political and social factors. This section describes some of these factors and shows how testing in the public sector differs from that of the private sector.

Specific federal, state, and local laws and strong public expectations combine to require extreme levels of openness, fairness and accountability in all aspects of testing and resulting personnel actions. The effects of this mandate are pervasive, beginning with public announcements of position openings, continuing with competitive testing of all applicants and limits on discretion in appointment/promotion decisions, and ending with appeals of any or all of these steps.

### **M**erit Principles

Most civil service jurisdictions operate under merit principles, either locally mandated or mandated by the federal government for programs receiving various types of federal funding, or both. These principles cover all personnel activities including selection. The merit principles, as they appear in the latest version of the Federal Standards For A Merit System Of Personnel Administration (U.S. OPM, 1983), are as follows:

- (a) Recruiting, selecting and advancing of employees on the basis of their relative ability, knowledge and skills including open consideration of qualified applicants for initial appointment.
- (b) Providing equitable and adequate compensation.
- (c) Training employees, as needed, to assure high quality performance.
- (d) Retaining employees on the basis of adequacy of their performance, correcting inadequate performance, and separating employees whose inadequate performance cannot be corrected.
- (e) Assuring fair treatment of all applicants and employees in all aspects of personnel administration without regard to political affiliation, race, color, national origin, sex, religious creed, age or handicap



and with proper regard for their privacy and constitutional rights as citizens. This "fair treatment" principle includes compliance with the federal equal employment opportunity and nondiscrimination laws.

- (f) Assuring that all employees are protected against coercion for partisan political purposes and are prohibited from using their official authority for the purpose of interfering with affecting the result of an election or a nomination for office.

These merit principles have considerable practical import for the day-to-day operation of a civil service selection program. Together with related civil service laws and rules they shape the civil service hiring process. Some of the major provisions are described here.

### **Public Announcement of Openings**

Most government jurisdictions have laws which mandate open announcement of all vacancies. Often the announcement takes the form of a one or several page flyer for each job title. (Selection is oriented to a class of jobs, all with the same job title, rather than individual jobs.) Typically the announcement must be displayed for a minimum period of time (three weeks in Massachusetts), and in certain specified public places (e.g., city halls, government offices, or newspapers). The announcement must state the content of the selection process, typically in terms of the test components, the areas to be tested, the weights to be assigned to each component, any entrance requirements (e.g., education or experience requirements), and, in some cases, any passing points which will be used in the examination. This openness forces specificity in assessment methods at an early stage in the testing process. There may be little room to tailor the assessment process to the number or quality of applicants, except based on past experience. This also often leads to large numbers of job applicants, placing further practical restrictions on the type of examination which may be employed.

## **Open Competition for Jobs**

Another common feature of civil service hiring is the requirement that all applicants be allowed to fairly compete for job openings. For example, the Connecticut civil service statute states that examinations:

“shall be competitive, free and ... open to all persons who may be lawfully appointed to any position in the class for which examinations are held...”

Practicality is not a consideration. Thousands of applicants may be tested for a mere handful of job openings. Since the form and content of the examination is typically announced in advance, this can result in very costly and time consuming tests with little value in terms of delivering government services. This is true whenever an examination mode is used in which the staff time invested is proportional to the number of applicants. For example, Oakland, California will administer an oral examination over a three month period to some 800 to 1,000 candidates for the position of fire fighter. This is part of a larger examination process, including a written and a physical ability test, which will screen 5,000 applicants for a total of perhaps 80 to 90 job openings over the next three years. The impracticality of such proportionate-labor examinations for most titles leads to the widespread reliance of government jurisdictions on the machine scored, written, multiple choice test for civil service examinations. With that form of examination, administration labor does not increase directly in proportion to the number of applicants.

## **Testing: Selection Based on Merit**

Merit selection is a cornerstone of all civil service systems. Typically the civil service legislation requires that persons who are appointed be shown to be capable of performing the primary and dominant duties of the position and assessment is restricted to those areas which can be measured reliably and fairly. For example, the Massachusetts civil service law states:

“Examinations shall fairly test the knowledge, skills and abilities which can be practically and reliably measured and which are actually required to perform the primary or dominant duties of the position for which the examination is held.”

Similarly, Connecticut law states that examinations:

“shall relate to matters as will fairly test and determine the qualifications, fitness and ability of the persons tested to perform the duties of the class or position to which they seek appointment.”

This type of legal mandate has led jurisdictions to focus on content-valid personnel selection tests which measure the capability of a person to perform the job. Typically the tests are designed to measure the knowledges, skills and abilities (KSAs) required to perform the job. Alternatively, the tests are designed to include samples of job duties themselves. Often, the test design considers both the KSAs and the important job duties. Omitted from the testing process are those areas which cannot be reliably and fairly tested, including various personality variables such as honesty, willingness to take risk, and willingness to assume authority. This leads to a very practical and basic orientation in the choice of areas to be tested. Typical test areas include reading comprehension; mathematical ability (of a type and level appropriate to the job); various specific job knowledges, such as knowledge of the principles and practices of criminal investigation (for a police promotional examination), or the knowledge of the principles and practices of filing (for a clerical position); and the ability to apply these knowledges.

Areas not tested usually include personality factors, personal characteristics, attitudes and preferences, motivation and those KSAs not amenable to testing with a multiple choice examination. Personality factors are typically not considered amenable to reliable and valid testing nor to have demonstrable relationships with job performance. Personal attitudes and preferences and motivation are typically and similarly omitted for two reasons. First, measures of these areas are so open to faking and subjective evaluation that they are not considered amenable to fair and reliable measurement. Second, these areas may not fit under the legal mandate to test the KSAs required to perform the job. The KSAs not amenable to measurement with a multiple choice examination might include honesty, creativity, courage and some aspects of the practical application of the KSAs which are measured.

The requirement for selection based on merit usually precludes random selection, an administratively simple and

inherently fair method of reducing a large number of applicants to a smaller, more manageable group for further consideration. There is also some tension between the concept of merit and that of affirmative action. Parenthetically, we note that there have always been non-merit factors considered in civil service hiring, the most pervasive and oldest of which is the government hiring preference given to war-time veterans.

### **Discretion in the Appointment Decision**

Typically appointments to civil service jobs are made from a ranked list of qualified applicants which resulted from the testing process. There is great variability in the amount of discretion given to the appointing authority. However, complete discretion is unusual. For example, in some jurisdictions an appointing authority with one opening to fill must appoint a person who is among the top three candidates on the list; for two openings, from among the top five candidates; and for three openings, from among the top seven candidates. This common method is known as the "Rule of Three" or the "Rule of  $2N+1$ " where N stands for the number of openings. There are variants even within the Rule of Three. For example, in Massachusetts an appointing authority making two appointments is required to appoint at least one of the top three candidates and cannot choose to appoint only the candidates ranked fourth and fifth.

Often, if the appointing authority chooses other than the top person on the list, some legitimate justification must be offered. This justification may be a matter of public record, open to inspection by any person and particularly by the persons not appointed. An appropriate justification may concern the ability of the candidates to communicate orally, as revealed in an employment interview.

### **Appeal Rights**

Depending on the jurisdiction, applicants may be given appeal rights at any of a number of points in the selection process. For example, in some jurisdictions the applicants have a right to review their test papers and the answer key which was used. For many years New York City carried this to an extreme; publishing each civil service examination after its use. Obviously either of these methods greatly reduces the ability of a jurisdiction to

refine and reuse a test. Some jurisdictions do protect the security of their examination materials by prohibiting post test review of either written tests alone or of any examination materials.

In many jurisdictions the applicants have the right to contest the form and content of the test if they feel the test is not a fair measure of their ability to perform the primary and dominant duties of the position. The applicants may have the right to appeal individual test items in addition to whole test components. The grading is also usually open to review or appeal. In the case of multiple choice tests, this may involve hand-scoring of one or more answer sheets. In the case of an essay examination, it may involve a far-reaching probe of the nature of the question and answer.

The propriety of the final appointment decision may also be appealed. An applicant high on the list may feel that the appointing authority gave an untrue or pretextual reason for choosing a lower placed applicant. The facts behind such decisions may be examined in detail before an independent body. These appeals are not *pro forma*. Often there is an independent government office set up to hear such appeals and empowered to direct remedies, including the firing of an applicant hired in violation of the civil service system.

These civil service appeal bodies may hear not only testing and selection appeals but also appeals of disciplinary actions, layoffs and other adverse actions. Individuals appointed as members of appeal panels are usually lay persons, often from two political parties and representing both labor and management. The standard for appeal varies among jurisdictions, but the orientation of the appeal body is usually very down-to-earth. Esoteric tests and testing theory are usually foreign to these groups. These conditions make for very conservative testing practice. Testing methods which rely on subjective judgment or modest levels of validity evidence are often shunned in favor of testing methods which will be easier to defend before the civil service appeal body.

## **Common Characteristics of Federal, State and Local Governments**

All levels of government (federal, state and local) employ civil service employees and most have formal civil service systems. These jurisdictions share a reliance on fair and open competition for jobs and share a system of division and checks and balances of power which affect personnel decision making. Some of the major common aspects of all three levels of government are described, followed by a short description of some of the unique aspects of each level of government as they relate to civil service employment.

### **Division of Power**

There is no one person who heads the government. Rather, there is a planned division of power. For example, at the state level the legislature mandates and funds the civil service hiring program. The governor then carries out this mandate as he/she sees fit, often embellishing or ignoring portions of the mandate. The final review of the immediate testing and hiring actions of the governor (and his/her appointees and the employees in the executive branch) is the responsibility of the courts. Any appeals not adequately resolved at an administrative level may be, and often are, heard in the courts.

Due to this division of power, it is quite difficult to effect change in the public sector. Even if a jurisdiction wishes merely to pilot test a new approach to personnel selection, it is probably necessary both to suspend some departmental rules and regulations on civil service hiring and to get specific permission from the legislature. Substantive change in civil service law probably entails a full-blown political process, involving public hearing and legislative action. And, in the process of lawmaking, logic and scientific merit are not the only relevant factors; many diverse political and social interests must be recognized. This often makes change very slow, no matter how sensible and practical the change appears to the civil service agency.

### **Openness**

All procedures and decisions are open to public scrutiny. The public sector hires in the public "fish bowl." This publicity makes it difficult to tailor the process to unforeseen developments (few

or many applicants, high or poor quality of applicants) as the process unfolds. Often not only applicants are monitoring the hiring process but also the news media. Due in part to the large number of applicants, the careful records on applications and appointments, and the openness of the system, class action employment-related lawsuits are common in public jurisdictions.

### **Time Required to Hire**

The stringent requirements of public announcement and open competition and the various administrative appeal rights often combine to create delay in the operation of the civil service hiring process. From the time a person applies for a job to the time of appointment may be many months. In Massachusetts, the examination announcement must be posted for three weeks. Then three more weeks are allowed before the examination for scheduling applicants and arranging examination sites. Grading takes some 30 days, which is followed by a 2 1/2 week period to allow appeal of aspects of the grading, after which a list of qualified applicants can be established. Then the list must be sent to the appointing authority who in turn interviews the candidates. Mailing the list may take two weeks, and scheduling and holding these interviews can easily take another three weeks, for a total of about 18 weeks. If the hiring process were to be delayed to resolve administrative appeals, the process could easily be extended another two weeks (for a simple review) to many months (for a full evidentiary hearing and decision). Thus, the time lag between date of application and date of hire is at least four to five months, even longer if delays occur.

From the appointing authority's viewpoint, this delay may be even greater. Prior to requesting a list of qualified applicants, the appointing authority must go through some process for approving the expenditure. Also, if a list of candidates is not in existence, the civil service agency may be required to develop a test, and this must be scheduled and implemented. This process might take a minimum of three months and may extend to many months and even years.

Although such extreme delays may not be the rule among civil service agencies, even a two-month time frame common at the local government level can cause criticism. Such delays in hiring discourage qualified applicants from applying and

completing the process, encourage appointing authorities to find ways to circumvent the civil service hiring system, and may even create a cadre of "provisional" employees, hired outside the civil service process for a temporary or limited time period. The existence of provisional employees creates a strong motive for agency managers to circumvent the mandated civil service system. In some jurisdictions provisional employees may serve for many years pending completion of a civil service hiring process and may comprise a sizable portion of the government work force.

### **Vagaries of Funding for Research**

The civil service examination process is carried out by a staff agency. In times of shrinking budgets, and arguably at all times, staff agencies are more likely to suffer budget cuts than line agencies, such as those delivering services to ill patients or those providing custodial services at correctional institutions. Within the civil service agency, the research and development (R&D) units which work to improve the selection process suffer a similar fate. They are seen as staff units by the civil service agency itself, as opposed to the line units which administer and grade the examinations. Despite nationwide pressures for improved selection methods and sometimes despite specific legal mandates for research into selection methods, R&D units are often minimally staffed, and this staffing is decreasing. For example, the R&D staff of the U.S. Office of Personnel Management has shrunk from 85 to 32 over the past 5 to 10 years. Many states and cities have no R&D function at all.

The small and shrinking funding for personnel selection R&D may be a reflection of the more general difficulty human resource professionals have convincing others of the dollars-and-cents impact of their programs on productivity and on overall organizational effectiveness. The need for a system to promote and support research and development is discussed in a later section of this paper.



## **Differences Among Federal, State and Local Civil Service Systems**

Although there are many commonalities, there are also differences among the assessment and hiring systems used at the several levels of government. This section describes some of the characteristics of the civil service systems which exist at each of three levels of government.

### **Federal Government**

The federal government is the largest employer in the United States. There were 3,133,303 federal civilian employees, as of July 1988, employed in several hundred different federal agencies and commissions located in all fifty states and in most countries abroad. Federal employees work in a very wide range of occupations, covering most clerical, technical, laboring, craft and professional jobs. While most federal employees work in the Executive branch, there are also large numbers in both the Judicial and Legislative branches. The immense size and complexity of the federal government presents personnel management issues faced by few employers.

The size and diversity of the federal work force along with the requirements pertaining to a public jurisdiction present major challenges in the areas of personnel selection and promotion. The federal government has devised a highly structured personnel management system as a way of maintaining order and control. There are classification and pay standards which assist in standardizing the structure of jobs and the salaries paid for various types of work across agencies and locations. There are qualification standards established to define the minimum qualifications for similar jobs across locations and agencies. There is a large Federal Personnel Manual which details policies and procedures and encompasses many volumes of text. In particular, Chapter 335 deals with policies and procedures for testing. These documents are designed to standardize procedures throughout the federal government. They are centrally developed by the U.S. Office of Personnel Management.

For generally common jobs which cross agencies, such as clerical and administrative, the U.S. Office of Personnel Management develops and administers the selection devices. These generally include written tests and ratings of training and

experience. The rating of training and experience is the most widely used selection and promotion device in the federal system. For the common occupations, candidates file applications and are tested at a central location. They are considered for job openings at a variety of locations and agencies as they occur. For occupations which are unique to particular agencies, the process is sometimes decentralized and administered by the particular agency. Selection for entry into competitive federal jobs is from the three highest scoring candidates by the agency selection official. Interviews usually form the basis for this selection.

Promotion decisions are made by individual federal agencies. Each agency develops procedures, which usually include a rating of training and experience and job performance appraisal. Written tests are rarely used as procedures to determine promotion to higher levels, but assessment centers, structured oral examinations, and work sample performance tests are occasionally used. For promotion testing, unions often become actively involved in the decision-making concerning weighting of test components, method of examination, passing points and other technical areas. The major concern of unions is rarely test validity. There is significant variation among agencies in the types of tests used for promotion. Some federal agencies have highly structured, centralized systems for the development of promotion procedures, while others are extremely decentralized, with each location developing its own procedures under some general guidance. Selection for promotion is usually determined by the selection official from a group of candidates who have been determined to be best qualified. (This may be from three to ten or more individuals.)

Because of its size and structured systems, the federal government's use of employment tests may appear to be complicated. The U.S. Office of Personnel Management works to insure the use of valid procedures by conducting research, providing training, developing standard tests and providing guidance on testing for the use of the large number of federal personnel professionals.

### **State and Local Government**

Most states and many cities and counties select employees based on merit, many through a formal program of civil service

examination and employment. This section gives some examples of the range of variation in civil service hiring systems found at the state and local levels of government.

The oldest state civil service system predates the federal civil service and rivals it in the complexity of the civil service law. The Massachusetts Civil Service is over 100 years old and has a civil service law which is 36,236 words in length, the equivalent of some 145 typed pages. (Conversely, Texas has no civil service system at the present time.)

Many civil service statutes are extremely long and convoluted due to many years of small changes, each addressing small aspects of a large system, and to many changes put in place by special interest groups.

There are wide variations among the states in the approaches to each of the basic components of a civil service system. The discretion allowed in making appointments from a civil service list or register varies greatly from state to state. For example, if there are 26 job openings, the appointing authority in Connecticut is able to choose from among any of the candidates who passed the examination. In Massachusetts, the choice would be limited to the top 53 candidates on the list. This is because Connecticut allows discretion among the top  $4+N$  ranks, while Massachusetts operates under a rule of  $2N+1$  candidates. Another distinction between states affecting this latitude is the source of legal authority. In some states the degree of latitude in appointment is stated in the civil service statute itself. In other states, it is the subject of a rule or policy. Still others have a basis in their state's constitution.

Some jurisdictions, such as the states of Washington and Massachusetts, have adopted a certification process known as "Three plus Three" as an affirmative action tool. This sometimes controversial policy allows the appointing authority to select from among the top three candidates and the top three candidates from among a specific protected class for one opening. In some jurisdictions, there must be a showing of "underutilization" or disparate employment rates for protected class members before this type of appointment process is allowed.

The number of preferred classes, and the nature of the preference, varies from state to state. Some states allow five to 10 points for veterans, other states give absolute preference to

veterans, at least in the examination score. Some states also give preference to diverse other groups, such as the sons and daughters of police officers killed in the line of duty or the mothers of veterans killed in wars.

An issue of great interest to applicants is the passing point for civil service examinations. In some states this is set at 70% of the questions asked in the examination (for at least some job titles). In other states it is left to the discretion of the civil service agency. Sometimes the test scores are scaled using a mathematical transformation which sets the passing score at 70 and distributes the other scores above and below that point. In other states such transformations are never done.

Some states are using a general entrance examination for selection into many entry-level professional positions. For example, Connecticut has developed the Connecticut Professional Trainee Examination which is used for entrance into many job titles, such as Social Service Eligibility Technician, Fiscal Administrative Officer and Accountant. Other states seem to have been influenced by or are following in the footsteps of the federal government which abandoned its Professional Administrative Careers Examination about ten years ago. For example, Massachusetts has no one examination for entry level positions. The decision to use one such examination is not usually made based on the validity or utility of the examination but rather on social and political grounds.

Some states, such as Massachusetts, use written examinations as the major mode of examination for promotion, others, such as New York State, for its technical and professional titles, use oral examinations as a major mode, others, such as Connecticut, use merit promotion boards of various types, while many others rely on structured evaluations of training and experience (T&E examinations) as the major mode of promotional examination.

Union involvement in the examination process varies widely. Massachusetts is now required to consult with labor about examination form and content. Connecticut municipalities have a stronger mandate to gain approval of labor for any changes in the municipal examination program. California has given even more power to its unions. On the other hand, Tennessee has no such

formal role of labor in the examination process as it has no state union.

For municipalities in some states, such as Ohio, Pennsylvania and Texas, the state legislature gives considerable direction to those administering the municipal civil services. A few states (e.g., Massachusetts) provide all, or virtually all, the civil service examinations for the municipalities which have civil service employees. In yet other states (e.g., New York, Connecticut) the state civil service agency will provide examinations for municipalities at their request, and sometimes only for a fee.

A few of the other areas where there is considerable variation in law and practice in state and local government are listed here without further comment:

- methods of ranking and certifying candidates (by score, or by band)
- the degree to which ranking of candidates is affected by and even mandated by veteran's preference laws
- amount of specificity in law, rule, policy and procedure
- affirmative action methods and goals
- general reputation and degree of public support for civil service
- consent decrees, court orders, threat of lawsuits
- size of staff dedicated to examination development
- size of staff dedicated to examination validation
- funding levels

Having discussed the variations in formal civil service systems, it must be noted that in many jurisdictions there are some positions which are exempt from civil service. In Massachusetts, approximately half of the employees of the executive branch are exempt from the civil service hiring process. Many of these are in lower paying, high-turnover positions. Thus there may be parallel and unequal systems in operation for similar or even the same job titles, even within one agency, but usually for different titles in different agencies. Also, in some jurisdictions there is yet another *de facto* dual system of employment. This is the long hiring process, and the cumbersome nature of the civil service employment system, which can result in a high proportion of persons hired "provisionally," that is, without benefit of a civil

service hiring process. Such provisional employees may serve for many years, completely outside the intended civil service employment system.

In conclusion, there are many differences as well as many commonalities in state and local systems of civil service employment. The bases for these features of civil service are often state and local statute.

## **C**omparison of Selection in the Public and Private Sectors

We can further clarify the context in which public sector personnel assessment operates by comparison with that of the private sector. As discussed above, the public sector operates under constraints not present in the private sector. A civil service agency has a responsibility to the public which goes beyond that seen in the private sector. This responsibility goes beyond positive labor management relations to fair and effective personnel selection and promotion.

Unlike the private sector, both the public at large and the applicants for public sector jobs expect and demand a selection process which is straightforward, logical, fair and open. Have you ever heard of an unsuccessful private sector applicant demanding a hearing before an impartial party to see if sound, fair hiring practices were followed? Some differences in assessment, selection and promotion between the public and private sectors described or alluded to in this paper are presented in the following table. The table on pages 18 and 19 gives examples of constraints and requirements in both the public and private sectors, and shows some of the additional constraints found in the public sector.

Private companies may agree with several or all of the admirable goals listed in the table, but they are not legally bound to adhere to all of them in every instance of hiring and promotion. There is the nub of the difference: the public sector is legally mandated to adhere to the highest standards in selection and promotion, the private sector does so whenever possible and practical.

These factors influencing and constraining personnel selection also shape research policy and practice. In both public and private organizations the selection of personnel has a major effect

on the productivity of the organization. However, this is more easily recognized in the private sector. In the public sector there is rarely a measurable "bottom line." As a result, the focus is on process rather than product. Thus the effects of poor personnel selection practices are difficult to recognize, and there is little objective, empirical support for funding needed to carry out sound personnel practices and the unavoidable R&D which this requires. Due, in large part, to the additional constraints on hiring, selection research in the public sector is more often motivated by a call for more timely and fair selection procedures. Personnel research in the private sector is more likely to focus on improving the productivity of the work force.

**Table 1**

**Some Factors Shaping and Constraining Personnel Selection In the Public and Private Sectors**

<b>Area</b>	<b>Private Sector</b>	<b>Public Sector</b>
<b>Goal</b>	Select good people	Allow all to apply and select the best among them
<b>Laws</b>	State/federal law on discrimination	State/federal law on discrimination State/federal/municipal law on merit and CS
<b>Ethics</b>	Corporate principles and policies	State ethics laws on conflict of interest and patronage Executive orders banning discrimination
<b>Fairness</b>	Fairness is an ideal goal Patronage is inefficient	Fairness is legally mandated Patronage is illegal
<b>Appeals And Reviews</b>	Internal review by personnel/AA office Review by Federal agencies	Internal review by agency personnel/AA office Review by Federal agencies  Appeals by individuals - to central CS office and/or CS Commission - to Human Rights Organization Routine audits by central CS
<b>Speed</b>	Speed in hiring possible	Speed difficult to achieve
<b>Power</b>	Centralized in CEO and Board of Directors Unions	Three equal branches of government Newsmedia Unions Special interest groups
<b>Ease of change</b>	Procedures and policies based on consensus or fiat Easy to innovate Change controlled by managers  Exceptions to procedures possible  Few legal ramifications or exceptions	Procedures and policies based in law  Difficult to innovate Change controlled by legislature and executive Exceptions to rules, policy and procedure are difficult and may be subject to appeal/review Exceptions may be grounds for appeal



<b>Area</b>	<b>Private Sector</b>	<b>Public Sector</b>
	Flexible Change to corporate policy or union contracts most difficult	Appeal at each step by persons involved Little flexibility
<b>Public interest</b>	Public has little interest in personnel matters	Public has great interest in personnel matters
<b>Central Control</b>	Degree of central control varies	Degree of central control varies
<b>Recruitment</b>	Area often local, may be countrywide Period may be arbitrarily short	Area usually jurisdictionwide, may be countrywide Recruitment period set by law, rule or regulation
<b>Paperwork</b>	Documentation desirable	Documentation required

**UGESP**

---

## *Special Conflicts In Public Sector Testing*

---

There are special conflicts in public sector testing that do not arise or are not as severe in private industry. These conflicts have arisen from the following major causes: conflicting statutory mandates, political mandates in conflict with statutory mandates, conflicting legal and professional standards, pressures from applicant groups in conflict with statutory mandates, and conflicting loyalties of human resource professionals. We will also discuss factors contributing to the conflicts.

### **C**onflicting Statutory Mandates

Both constitutional provisions and federal, state and local laws govern personnel selection and particularly the merit-based personnel selection of civil service work forces. These mandates do not all converge in a simple, synergistic fashion.

There is a critical conflict between statutes mandating test validity and fairness to individuals on the one hand and elimination of discrimination against minority groups on the other. Although these provisions are not inconsistent on the surface, since "adverse impact" is not discriminatory if the test is valid, the common phenomenon of valid tests producing adverse impact causes conflict. (In this discussion, adverse impact refers to situations in which protected group members have relatively lower test scores than majority group members.)

Employers are subject to Title VII of the 1964 Civil Rights Act, as amended, which prohibits employment discrimination on the basis of race, color, religion, sex, or national origin. Title VII is enforced by the Equal Employment Opportunity Commission and by the Office of Federal Contract Compliance Programs (OFCCP) using the *Uniform Guidelines on Employee Selection Procedures (Guidelines)*. Discrimination is also prohibited by various state and local fair employment laws. These laws are aimed at preventing discrimination in selection and promotion on the basis of group membership such as race or sex and require the employer to demonstrate business necessity or validity of the selection device when there is adverse impact against minorities.

While this requirement applies to all employers, public jurisdictions are subject to validation and fairness (to individuals) requirements whether adverse impact occurs or not. These requirements arise from the U.S. Constitution and federal, state and local laws, including:

- The Fifth Amendment to the U.S. Constitution, applying to the federal government, and due process clause of the Fourteenth Amendment to the Constitution, applying to state and local government.
- 42 U.S.C. 1983, which prohibits state and local governments from violating the Constitution or laws of the United States.
- State constitutions which often include language similar to that of the Fourteenth Amendment, applying to both state and local jurisdictions.
- The Civil Service Reform Act of 1978 (applying to the federal government), which provides for minority recruitment but specifies that "... selection and advancement should be determined solely on the basis of relative ability, knowledge, and skills, after fair and open competition which assures that all receive equal opportunity" (5 U.S.C., 2301b).
- State and local Civil Service laws with provisions which specify that employee selection and promotion will be based on job-related tests.

Based upon these Constitutional provisions and laws, the selection devices of government agencies can be challenged upon grounds of being arbitrary and irrational regardless of whether adverse impact exists.

The critical issue for public employers is what to do when valid tests have adverse impact. This is more troublesome for public employers; private employers do not have to face this issue since they are under no obligation to use valid selection devices if they can avoid adverse impact. Notwithstanding the U.S. Supreme Court's deemphasis of the bottom line in *Connecticut v. Teal* (1982) and its message regarding the illegality of quotas in *Watson v. Republic National Bank of Fort Worth* (1988), private employers can simply hire enough minorities to ensure that there is no adverse impact. And, regardless of the types of selection

devices they use, they have the flexibility of using them in ways to avoid adverse impact.

The solution to this dilemma in the public sector has been the use of affirmative action plans which involve consideration of the top-scoring minorities as well as top-scoring majorities for each vacancy. In this situation, race is a factor along with test score and other factors in final employee selection decisions. However, court decisions have necessitated careful balancing between the objectives of merit selection and fairness to individuals on the one hand and reducing adverse impact and increased hiring of minorities on the other. These court decisions have established the following points. (This discussion does not include *Wards Cove Packing Co. v. Atonio* (1989) since the U.S. Supreme Court decision had not been released at the time this paper was written.)

First, racial quotas are not acceptable. In their ruling in *Regents of University of California v. Bakke* (1978), the Supreme Court found that use of quotas preventing access to positions (in this case in the medical school) solely on the basis of race without regard to qualifications violated individual rights guaranteed by the Fourteenth Amendment.

This judgment was extended in the *Teal* decision which clarified that the 1964 Civil Rights Act, as amended, ensured fairness to individuals rather than to groups. At issue in this case was whether discrimination existed if an initial screening device produced adverse impact against minorities while the final selection decisions did not. Stating that failing black applicants could be expected to get no satisfaction from knowing that fellow blacks had been hired, the Court ruled that test validity must be demonstrated by the employer.

A third finding made by the courts which has drastically affected affirmative action plans is that population statistics as a basis for affirmative action plans are not acceptable. In *Janowiak v. the Corporate City of South Bend* (1987), on remand from the Supreme Court, the U.S. Court of Appeals, Seventh Circuit, determined that selection including race as a factor must be based upon an affirmative action plan demonstrating a manifest imbalance between percentages of minorities in the relevant qualified labor market and the work force. This decision

was based on Supreme Court decisions in the cases of *Johnson v. Transportation Agency of Santa Clara County* (1987) and *Wygant v. Jackson Board of Education* (1986). Although the *Johnson* decision was supportive of affirmative action plans in that an employer need not prove past discrimination to justify a voluntary affirmative action program, it was also specific in requiring the need to show a manifest imbalance between the relevant labor market and the work force.

And although applying to a layoff plan, *Wygant v. Jackson* is also critical to affirmative action based selection decisions because of its majority agreement on the following points, summarized by the U.S. Court of Appeals, Seventh Circuit, in *Janowiak v. City of South Bend*: "(1)... the plan must be justified by a compelling governmental interest and (2) the means chosen by the government must be narrowly tailored to effectuate the plan's purpose." Further, "...before a governmental unit can establish a compelling governmental interest in remedying discrimination, it must make some showing of prior discrimination on its part" (*Wygant v. Jackson*). This in turn was defined as a difference between percentages of minorities on its staff and in the relevant labor market.

These three cases begin to lay out the conditions under which affirmative action plans will be upheld and clarify that plans based on population statistics do not demonstrate the existence of past discrimination since they do not show a manifest imbalance of racial percentages between the relevant labor market and the work force. Therefore, although use of race as a factor in selection decisions along with test score has been the solution to the quandary of what to do when valid tests produce adverse impact against minorities, it is clear that this must be done prudently to ensure fairness to individuals.

Until the resolution of these two mandates is clarified, jurisdictions will have to deliberate over their application whenever merit selection results in adverse impact.

**Veterans Preference.** While conflict between merit selection and discrimination has created more tension, laws giving preference to veterans produce another source of conflict not encountered in the private sector. Usually providing for the addition of points to the test scores of veterans, these laws conflict with laws requiring validity and fairness since they reduce the

validity of the selection decision by adding a non-job-related component to the test score. In some jurisdictions this is becoming a minor issue because this preference is limited to war-time veterans, a group which is shrinking in number.

## **C**onflicting Statutory and Political Mandates

Even more difficult to resolve than such conflicts between statutory mandates are opposing philosophies behind conflicting legal and political mandates. Selection is often the focus of strain between political mandates and statutory mandates regarding fairness and merit selection.

### **Mandates for Affirmative Action**

Conflicts regarding fairness occur when legislative bodies and governors/mayors representing minority groups advocate affirmative action concerns to the extent that they are in conflict with the law. This occurs when such political bodies or leaders: set racial quotas instead of goals, a practice which was struck down by the Supreme Court in *Bakke*; adopt affirmative action practices which provide fairness to groups instead of individuals, a practice which was struck down by the Supreme Court in *Teal*; and set goals based on population versus labor market statistics, a practice which was struck down by the Supreme Court in *Wygant* and by the Seventh Circuit in *Janowiak*.

### **Political Mandates Regarding other Political Issues**

Apart from this problem, political mandates may conflict with merit selection mandates not due to affirmative action concerns but to old-fashioned political concerns. Political mandates may conflict with the traditional and mandated selection goal of choosing the best qualified person for the job. Quite apart from wishing to hire trusted political advisors as part of their management team, this occurs because newly elected politicians feel that they have a mandate from the public to carry out promised programs. In this situation, the most important characteristics of key subordinates are loyalty and commitment to political agendas rather than managerial skill and knowledge. This obviously conflicts with the merit principle of choosing the best person for the job based on knowledge and skill, which is required by federal and often state and local law. Although top positions can often be

exempted from civil service systems, the requirement of loyalty can come into conflict with merit principles at mid-management levels in an organization. While private sector organizations may encounter the same type of conflict, especially when there is a change in top level management, merit selection is not required unless adverse impact occurs.

## **C**onflicting Legal and Professional Standards

In order to comply with federal requirements and meet professional standards for the development and use of tests, assessment professionals follow the *Uniform Guidelines on Employee Selection Procedures* (*Guidelines*, 1978), the *Standards for Educational and Psychological Testing* (*Standards*, 1985), and the *Principles for the Validation and Use of Personnel Selection Procedures* (*Principles*, 1987). However, there are inconsistencies among these documents, particularly between the *Guidelines* on the one hand and the *Standards* and *Principles* on the other. Both the *Standards* and *Principles* are more recent than the *Guidelines* and are based upon relevant and recent research. Each is intended to constitute, as far as possible, a consensus of professional opinion, whereas the *Guidelines* form a legal document used to prove or disprove discrimination. Areas in which serious inconsistencies among the documents exist are: the rationale for setting cutoff scores, the use of tests for ranking versus screening decisions, the need for differential prediction studies, the use of content versus construct validation strategies, and the support given to validity generalization.

### **Rationale for Setting Cutoff Scores**

Cutoff scores or passing points determine which candidates are included on employment lists. Firm cutoff scores are more likely to be used in the public than the private sector. In their discussion of the evidence necessary to support screening decisions, the documents differ markedly in the degree of rationale required. Although allowing for setting a cutoff score based on the number of job openings and candidates at higher score levels, the *Guidelines* demand a detailed rationale for setting passing points based on normal acceptable proficiency: "Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable

proficiency within the work force” (p. 38298). And, “the user should describe the way in which normal expectations of proficiency within the work force were determined and the way in which the cutoff score was determined” (p. 38305).

The *Principles* on the other hand support any cutoff score if the test is valid: “Cutoff or other critical scores may be set as high or as low as the purposes of the organization require, if they are based on valid predictors... Judgment is necessary in setting any critical or cutoff score. A fully defensible empirical basis for setting a critical score is seldom, if ever, available. The only justification that can be demanded is that critical scores be determined on the basis of a rationale which may include such factors as estimated cost-benefit ratio, number of openings and selection ratio, success ratio, social policies of the organization, or judgments as to required knowledge, skill, or ability on the job” (p. 32-33).

The *Standards* require explanation of the method and rationale for setting cutoff scores as well as subject matter expert qualifications and determination of standard error of measurement at the cutoff score level. However, like the *Principles*, they do not demand an explanation based on level of proficiency.

### **Use of Tests for Ranking Versus Screening Decisions**

A second serious difference among the documents concerns the issue of using tests as screening or ranking devices. Based on potentially greater adverse impact due to ranking, the *Guidelines* require more validity evidence to support ranking decisions than to support screening decisions. “Evidence which may be sufficient to support the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the same procedure on a ranking basis under these guidelines” (p. 38299). And, “Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance” (p. 38303).

The *Principles*, on the other hand, assert that content valid tests are suitable for use as ranking devices. “In usual circumstances, the relationship between a predictor and a criterion may be assumed to be linear. Consequently, selecting from the top scorers on down is almost always the most beneficial



procedure from the standpoint of an organization if there is an appropriate amount of variance in the predictor. Selection techniques developed by content-oriented procedures and discriminating adequately within the range of interest can be assumed to have a linear relationship to job behavior. Consequently, ranking on the basis of such scores is appropriate" (p. 32).

### **Need for Studies of Differential Prediction**

The need for doing studies of differential prediction constitutes a third issue on which the documents differ. The *Guidelines* stress the necessity of investigating fairness when feasible: "We recognize that there is serious debate on the question of test fairness; however, ... we have retained in the guidelines the obligation upon users to investigate test fairness where it is technically feasible to do so" (p. 38294-5). The *Standards* also support conducting differential prediction studies with support designated as conditional or varying with the application. Standard 1.20 states: "Investigations of criterion-related validity for tests used in selection decisions should include, where feasible, a study of the magnitude of predictive bias due to differential prediction for those groups for which previous research has established a substantial prior probability of differential prediction for the particular kind of test in question" (p. 17).

While not directly addressing whether differential prediction studies should be done, the *Principles* conclude that its existence is not supported by the literature. Citing Schmidt, Pearlman & Hunter (1980), Hunter, Schmidt & Rauschenberger (1984), and the National Academy of Sciences (1982), they state: "There is little evidence to suggest that there is differential prediction for the sexes, and the literature indicates that differential prediction on the basis of cognitive tests is not supported for the major ethnic groups. There is no compelling research literature or theory to suggest that cognitive tests should be used differently for different groups" (p. 18).

### **Appropriateness of Content Versus Construct Validity Strategies**

The fourth major difference among the documents concerns the issue of when a content validity strategy will suffice to defend

use of a selection test and when a showing of construct validity is necessary. The *Guidelines* state that: "A content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated" (p. 38299). Also, "A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined ... and is a necessary prerequisite to successful job performance" (p. 38303). The *Guidelines* distinguish between mental processes, which cannot be adequately supported by content validity evidence, and KSAs, which can.

The *Principles* support use of a content-oriented strategy under narrower conditions: "to situations in which a job domain is defined through job analysis by identifying important tasks, behaviors, or knowledge and the test (or criterion) is a representative sample of tasks, behaviors, or knowledge drawn from that domain" (p. 19). This does not include situations "in which more general worker specifications (such as general skills or abilities) are measured and match well those inferred from the job domain" (p. 19). According to the *Principles*, construct validity evidence would be required in the latter case. The *Standards* also emphasize the necessity of the test's being a representative sample of the job. Standard 10.5 states: "When the content-related validation evidence is to stand as support for the use of a test in selection or promotion, a close link between test content and job content should be demonstrated" (p. 61). Also, "In general, content-related evidence demonstrates the degree to which the sample of items... on a test are representative of some defined universe or domain of content" (p. 10).

## **Support of Validity Generalization**<sup>1</sup>

The documents also differ in their degree of support for practices based on the evidence for validity generalization. While the *Guidelines* allow using validity evidence from other sources under limited conditions, the *Principles* and *Standards* make stronger statements about the appropriateness of validity generalization. According to the *Standards*, "Employers should not be precluded from using a test if it can be demonstrated that the test has generated a significant record of validity in similar job settings for highly similar people or that it is otherwise appropriate to generalize from other applications" (p. 59). The *Principles* eliminate the restrictions regarding similar job settings for highly similar people. According to the *Principles*, "To the extent that validity generalization evidence is available, researchers may rely on it to support the use of selection instruments" (p. 27).

## **Effect of Conflicting Standards**

As stated earlier, assessment professionals use these documents to comply with both federal guidelines and criteria of sound professional practice. Conflict among them creates dilemmas in making selection decisions, especially when the *Guidelines* demand evidence which does not comport with professional opinion or is infeasible and unnecessary. The difficulty in resolving these dilemmas is increased since little leeway is allowed in following the *Guidelines*. While the *Principles* and *Standards* present themselves as ideals not expected to be always achieved, the *Guidelines* in general allow no leeway for professional judgment and are often used in adversarial proceedings as the absolute standard of practice.

---

<sup>1</sup>Since this paper was written, a committee on the General Aptitude Test Battery convened by the National Research Council (Hartigan and Wigdor, 1989) has issued a report supporting validity generalization.

Due to these conflicts and the quantity and breadth of the professional standards, assessment specialists sometimes conclude that they do not have the time and budgetary resources to fully address all of the issues. Finding the optimum level of effort to put into the development of each test is problematic. One important test may use most of the year's resources of an organization, leaving little for other work. Consequently, they make judgments to focus on some standards more than others. Particularly in the public sector, these judgments are always open to challenge from other assessment specialists or from lawyers. This leaves public sector assessment open to complex and continuing legal challenges.

### **P**ressures from Applicant Groups

A fourth major source of conflict stems from pressures from applicant groups, whose philosophies often run counter to selection decisions based on validity and adverse impact concerns. These can be grouped as follows:

1. Professionalism versus adverse impact and the need for proof of validity. Occupations trying to establish themselves as professionals advocate requiring bachelors or advanced degrees as entry requirements. However, a degree requirement often increases adverse impact against minorities, and its validity is difficult to prove. Since the Supreme Court ruled in *Griggs v. Duke Power Co.* (1971) that use of an unsubstantiated requirement was a pretext for discrimination, such requirements have typically been reduced rather than augmented.

2. Seniority versus adverse impact and the need for proof of validity. Union groups traditionally advocate allocating jobs on the basis of seniority. However, this often causes adverse impact due to past effects of discrimination, and the validity of the length of employment is hard to prove. In fact there is evidence that length of experience is not a valid predictor of job success. (However, Title VII of the 1964 Civil Rights Act, as amended, specifically exempts bona fide seniority systems from the coverage of the act.)

3. Employee group concerns versus adverse impact and need for validity. Apart from unionized workers, employee groups in general advocate promotion from within the organization versus hiring from outside on a competitive basis, and hiring from the outside often impacts negatively on morale. However, if minorities have not been hired in the past, internal promotion causes adverse impact.

4. Handicapped and disabled group concerns. Handicapped and disabled groups challenge job requirements as unnecessary for performing duties and even challenge job duties as unessential to the job.

The wishes of these groups conflict with both adverse impact and validity concerns.

### **C**onflicting Loyalties of Human Resource Professionals

The fifth and final major source of conflict in public sector testing stems from the conflict in loyalties demanded of human resource professionals. While human resource departments in private organizations may have conflicting goals regarding efficiency and validity on the one hand and affirmative action on the other, at least they clearly report to one chief executive officer who can decide how to balance the two concerns. In public organizations, personnel staff must contend with conflicting loyalties. They are often directly responsible to the chief executive. However, they may also serve as a staff agency to a civil service commission or independent board designated to review personnel matters and function as a staff agency to line departments, many of which have department heads with independent authority as elected officials. They must be responsive to elected representatives, who are concerned with the complaints of citizen groups or individual citizens, and to independent citizens directly. Finally, professionals have a loyalty to the standards and ethics of their profession. For assessment specialists this frequently means balancing a respect for the *Principles and Standards* with the constraints of a fixed, limited budget for civil service examining.

## **F**actors Contributing to Conflict

There are several factors impinging on public sector selection that contribute to the conflicts discussed above. These factors affect selection decisions in the areas of recruitment, setting job requirements, and the use of tests as ranking or screening devices. Each is discussed below:

### **Recruitment**

Recruitment factors influence the degree of adverse impact in the selection process. It is easier for private industry to avoid adverse impact by recruiting the most qualified minorities. While private organizations can do whatever is necessary to recruit well-qualified minorities, public organizations are often restricted from effectiveness by the following:

1. Limitations on salary and fringe benefits. There is usually a defined salary range for each position, with only small leeway for maneuvering with the starting salary, and fringe benefits, even for management positions, are usually not subject to negotiation. Variations in the hiring offer cannot often be made quickly but may have to be approved by the legislative body as well as the executive branch.
2. Commitment to early promotion and salary raises. Early promotion can rarely be promised due to the probability of competition, and salary raises are limited due to public interest in keeping taxes down.
3. Flexible staffing patterns. Public organizations usually have little flexibility in staffing patterns. Private sector organizations can hire people in anticipation of vacancies. Since public sector hiring must be based on actual vacancies rather than potential need, recruitment cannot necessarily be done at the best times for finding the best candidates. This makes it particularly difficult to recruit and hire college seniors.
4. Advertising and travel budgets. Advertising and travel budgets are usually small compared to those of private industry and cannot easily be supplemented.
5. Competition after recruitment. Hiring decisions cannot be made in isolation; even specially recruited candidates must

compete with all others who are interested. If these candidates perform poorly on the selection test and/or are not hired, the credibility of the organization is negatively affected and ill-will is generated.

While private industry can act decisively to secure the best qualified minorities to avoid adverse impact, their public sector counterparts, who may be competing with them for the same candidates, cannot. In addition, aside from the difficulty these factors present in the recruitment of minorities, outside recruitment poses a difficulty in itself since it often conflicts with the interest of other groups. In addition to unions and other employee groups discussed earlier, which advocate promotional opportunities for their own members, these include hiring departments, which wish to fill vacancies as quickly as possible and may have some loyalty to employee groups, and local politicians, who advocate consideration for their constituents and are more interested in employment of local minorities than in broad recruitment for those best qualified. This interest is strengthened because of the widely-held belief that the public sector has some responsibility for providing jobs to its citizens, especially if they have been laid-off from private sector jobs.

### **Job Requirements**

It is relatively easy for private employers to adjust job requirements if necessary to achieve affirmative action goals. In public jurisdictions, however, there are pressures from the following groups (discussed in part earlier) to keep strong minimum requirements: professional groups wishing to upgrade their status by establishing higher educational requirements; employee unions pressing for promotion by seniority rather than through job-related tests; and departments advocating higher requirements to ensure on-the-job efficiency, to reduce the need for training, and to ensure public health and safety. Public health and safety issues are especially important for police, fire, building inspection, and health occupations. Public jurisdictions also often have residency requirements limiting job applicants to individuals who are residents of the jurisdiction. On the other hand, handicapped groups are challenging employment standards in general as well as the duties established for specific jobs.

With the exception of the need for a clear professional knowledge base such as for engineers, accountants, and physicians, it is difficult to conclusively demonstrate the business necessity of many job requirements. Elimination of job requirements, however, presents another problem. Usually the absence of job requirements produces more applicants who do not possess the requisite KSAs for the job. When a test is used under these circumstances, the entire burden shifts to the employment test to assess KSAs necessary to do the job and eliminate candidates not possessing these KSAs. If the proportion of unqualified minorities applying is high, adverse impact previously due to both requirements and the test will be shifted entirely to the test, potentially producing discrimination litigation.

### **Use of Test Results for Ranking Versus Pass-fail Decisions**

Public jurisdictions usually select employees based on test ranks but may also establish minimum cutoff scores. To reduce adverse impact, affirmative action advocates often urge abolishment of ranks in favor of selection from among any who have passed. As stated earlier in this paper, this position is supported by the *Guidelines*, which suggest that the standards for justifying the use of a test for ranking are higher than those for justifying a passing point. However, elimination of ranking also eliminates meaningful information regarding differences among candidates on KSAs critical to the job. Professional opinion, as reflected in the *Principles*, holds that ranking is more useful in serving the purposes of the organization, even when no empirical evidence exists in the form of criterion related validity.

Advocates for the elimination of ranking presumably believe that a passing point can be justified based on evidence of normal acceptable proficiency as suggested in the *Guidelines*. However, unless the test has been developed according to criterion-referenced measurement principles, the justification of a cutoff score depends on the validity of the test. Moreover, although any cutoff score can be justified if the test is valid, it is difficult to defend the choice of any particular cutoff score to differentiate between qualified and unqualified candidates.

The theory underlying a pass-fail use of a test to differentiate between qualified and unqualified candidates has been largely



discredited. To justify a cutoff score the relationship between test score and job performance would have to be a step function, with low or unsatisfactory job performance predicted by scores below a certain point and high job performance predicted by test scores above that point. This is not found in practice (Anastasi, 1988, pp. 167-168). Rather the relationship is linear, with no discrete breaks in the relationship between test and job performance. To describe the profession's view, both the *Principles* and the *Standards* support setting cutoff scores based on such factors as the selection ratio, number of openings, and cost-benefit ratio.

A third difficulty with the use of cutoff scores only is that hiring departments would then presumably be obligated to consider for employment all candidates who passed the test. Depending on the number of candidates involved, selecting from among them would prove extremely burdensome.

## Conclusion

There are many legal, political, social and professional views and mandates concerning public sector personnel assessment and selection. Although these coexist, many are or have the potential to be in conflict with one another. Due, in part, to the public and tripartite nature of our governments, these conflicts have not yet been resolved. Personnel assessment specialists are frequently caught in the middle, forced in almost an existential fashion to resolve these conflicts in a given matter by the actions they take.

---

## ***Constraints On The Use of Testing Methods In The Public Sector***

---

Because of the unique perspectives of the public sector which have been outlined in this paper, there are special constraints on the feasibility of various types of tests and the use of those tests. Since most public jurisdictions function under a merit system, applicants have many legal rights. They are usually able to question or appeal the testing process, and employers must be able to explain why a particular type of test was used and how it was scored. Because of the openness of the environment in which personnel testing takes place, employers are concerned that applicants accept the testing process. In addition, applicants expect their rights to privacy will be respected.

Applicants expect testing procedures which are fair and equitable to all. Public jurisdictions usually actively recruit applicants and often must test large numbers of people who are competing for a relatively small number of job vacancies. In order to ensure fairness to all applicants, public jurisdictions usually provide tight security for the test materials to insure that no one has had benefit of inside information about the test. This can present problems for reuse of test materials. The large numbers of candidates often make certain types of tests infeasible.

As discussed above, tests used by public sector employers must meet the same stringent technical and legal requirements as tests for any other employers. Tests used must be valid and reliable. Public employers are covered by Title VII of the 1964 Civil Rights Act, as amended, and must be able to demonstrate the validation of tests if the use of those tests results in an adverse impact against a protected group. In addition to federal laws covering the use of tests, state and local employers often are subject to state or local laws or regulation concerning personnel testing. Often the local statutes are specific and detailed in their requirements concerning the nature and use of testing instruments.

## **T**est Techniques

The public sector has a number of testing techniques in common with the private sector. In addition, the public sector has techniques which are unique or used differently. In general, in order to be able to demonstrate objectivity and fairness to all candidates, tests used by the public sector tend to be highly structured. As noted, they are used to pass or fail candidates and often to rank order them in terms of each candidate's relative fitness for the job. The paper prepared for the National Commission on Testing and Public Policy by Reilly and Warech (1988) discusses the general research concerning a number of the test modes mentioned here. In addition, the paper by Sproule (1989) summarizes work on a number of these tests in substantial detail, and a limited-distribution publication edited by Wright (1974) and one authored by Maslow (1983) describe public sector perspectives on test methods. This paper will highlight the unique perspective public employers have in the selection of testing techniques.

### **Rating of Training and Experience**

A common form of examination in the public sector is a rating of training and experience (T&E). A T&E lays out a specific rating process by which each candidate's background is evaluated. The rating process allows the employer to demonstrate to candidates how their backgrounds are evaluated and the criteria against which all applicants are evaluated. This method of examination is one of the most widely used in the public sector. There are a variety of approaches to these ratings. A T&E is usually based on premises such as the more directly related or greater the amount of experience or education applicants have, the better they will perform a job. Traditionally, T&Es rarely attempt to assess the quality of past experience or education. The several T&E methods are based on assumptions similar to those private employers make when they base selection on a review of resumes. The difference is that T&Es provide a more structured evaluation of the data; that is, they are scored like tests.

In its most primitive form, a T&E rating awards points for each year of experience or education beyond the minimum qualifications. In such a basic T&E the information rated would be contained in the applicant's standard application form. There

is limited evidence of validity for this type of T&E (McDaniel *et al.*, 1988).

There are a number of more sophisticated methods of ratings of training and experience which have been developed in recent years. These methods usually rely on a structured questionnaire to gather information for the ratings. Structured questionnaires are developed for the specific jobs and gather information directly relevant to the job being filled, rather than relying on general information presented in a standard application form.

Among the newer T&E methods is one method which awards points for various types of experience as they relate to required areas of knowledge or skill. The areas of knowledge, ability and skill are identified through job analysis, and training and experience ratings are linked to the areas. This method follows a content validation model. A similar method is referred to as the grouping method. Rather than awarding points for job-related experience and training, this method defines categories of training and experience. The candidates are assigned a score based on the judged category of their experience and training. In the self-rating method of T&E, candidates respond to highly structured lists of job tasks or knowledges, skills and abilities and indicate the level of their experience or training as it relates to each specific task or KSA. Finally, the most highly researched and most promising method of training and experience evaluation for higher level jobs is the behavioral consistency method. This method relies on the evaluation of job-related accomplishments described by candidates on a structured questionnaire.

Public sector employers often use T&Es as both the basis for selection of employees and also their promotion to higher level positions. They are generally viewed as efficient methods to test relatively large numbers of applicants and are generally accepted by both selecting officials and candidates. Research in Connecticut (Wiesen, 1988) indicates that candidates prefer this method of examination to both written and oral examinations.

A limited amount of research has been conducted on the validity and reliability of the various approaches to T&E examination. Ash and Levine (1985) compared the validity of four methods, while McDaniel *et al.* (1988) conducted a meta-analysis of T&E methods. In both cases, the research indicated that traditional methods tended to have little or no validity while

newer methods, especially the behavioral consistency method, showed higher validity. These are the major studies concerning the validity of these widely used tests. Little research, however, is available about the adverse impact of this testing method. Certainly there is a great need for additional research, especially on the relative validity, reliability, adverse impact, fairness, and utility of various approaches to rating training and experience.

### **Merit Promotion Boards**

The merit promotion board is a method used to promote employees to higher level positions. This method is commonly used by federal agencies to promote employees. In this testing method, applicant personnel files are reviewed using a training and experience evaluation, usually combined with other information concerning candidates, such as job performance evaluation, awards, or a panel interview. This method is similar to the judgments made by private employers concerning whom to promote but employs a structured, scored rating process, clearly defined so that its fairness is evident to all. Unions often play an active role on these boards as participants or observers.

### **Structured Oral Examinations**

Many public sector jurisdictions (e.g., New York State) use structured oral examinations (oral boards), particularly for selection for promotion to higher level jobs. A structured oral examination consists of a set of questions and, often, follow-up questions which have been developed to measure important abilities and knowledges identified by a job analysis. The examination is usually conducted by a test administrator, but the questions are asked and answers rated by a panel, usually comprised of several experts in the subject matter. The questions vary in content. Structured oral questions may be factual knowledge questions, hypothetical situational questions, questions based on past applicant experience, or even role playing exercises. An important characteristic of these tests is that the questions have been developed prior to test administration and are asked in a standardized way to all applicants. As in the case of the other tests discussed, public employers must be able to demonstrate the fairness, objectivity and standardization of the process.

Another important characteristic of structured oral examinations concerns the development of rating criteria. Criteria often are developed concurrently with the questions and are based on a job analysis. Raters or oral board members usually receive training on rating criteria prior to test administration. Candidates receive scores based on the rating criteria and are usually rank ordered based on these scores.

Recent research has indicated that structured oral tests are a valid technique for testing (Campion *et al.*, 1988, Silverman and Wexley, 1987, and McDaniel *et al.*, 1987). This is welcome news to public sector agencies which have used this technique for many years as the basis for selecting individuals for many high level, responsible positions in government. In addition, this method may result in less adverse impact than written tests.

### **Multiple Choice Tests**

The multiple choice written test for many years has represented "civil service testing" to the general public. This is because the written multiple choice test is widely used, particularly for jobs which attract large numbers of applicants. Aptitude or ability tests are often used as the selection instrument for entry-level jobs because many jobs have no job knowledge prerequisites. Job knowledge tests are frequently used as the basis for promoting employees to higher level jobs. Written tests are commonly used to examine candidates for police officer and firefighter jobs, are highly visible to the general public and have been the subject for numerous legal challenges of employment discrimination.

Written multiple choice tests, however, continue to be used widely by public sector jurisdictions. Because of their ease of administration and scoring, they are an extremely efficient method for examining large numbers of applicants. They are also easy to explain to candidates, in that only one answer is correct and all candidates are scored in the same objective manner. In some jurisdictions, hearings are even held to finalize the answer key, resulting in scoring criteria that are clear and defensible. Considerable effort has been expended to assure that these tests are free from bias (e.g., Ashton & Wiesen, 1983; Flaughner, Nieves, Slaughter, Wiesen and Woodford, 1980; Wiesen, Reynolds, & Estes, 1979). Meta-analytic studies have shown that written multiple choice tests are valid measures for a wide range of jobs

(Hunter and Hunter, 1984, Schmidt and Hunter 1981, Schmitt *et al.*, 1984). In particular, job knowledge tests are among the most valid tests available, if not generally the most valid.

While private sector employers use written multiple choice tests as part of the selection process, public agencies differ in their use in several important ways. First, written multiple choice tests are used more extensively to test large numbers of applicants for large numbers of jobs. Second, candidates must receive passing scores to be placed on the employment list, and they are rank-ordered based on the scores. Third, written multiple choice knowledge tests are used as the basis for making decisions for the promotion of employees to higher level jobs. And fourth, public sector applicants are given their scores.

While written multiple choice tests continue to be used widely by public agencies for personnel testing, they often result in an adverse impact against minority members. This continues to be an area of concern, in particular, for public employers.

### **Other Written Test Formats**

Besides using written multiple choice tests, public sector jurisdictions use other formats for written tests. These include essay tests and written simulations.

Essay tests are highly structured. They use standard questions and structured rating criteria for grading, often as a method for assessing written communication skills.

In a written simulation test, candidates are presented with a situation similar to that which might be faced on the job. They are presented with alternate courses of action. Depending upon their choices, candidates are given further events and continue to select choices and be directed until the situation logically ends. These problems reflect the idea that there is not always one clear-cut best course of action. Another type of written simulation involves the use of video tapes as stimulus materials to which candidates must respond using a written multiple choice format. Still another type is the latent image test, a relatively new development in employment testing. These tests are not widely used at present primarily because of the complexity of the test development process. These tests are a variation of situational multiple choice written tests but attempt to more closely simulate reality. Little research has been done to date on the validity,

reliability and adverse impact of these tests. But from a content validation viewpoint, they represent an attempt to more closely simulate work behaviors while retaining the practical benefits of written multiple choice tests.

### **Performance Tests**

For certain types of jobs, particularly skilled blue collar, craft and clerical jobs, public sector jurisdictions have relied on performance tests. These tests usually require the test taker to perform sample work behaviors of the job. As in other test methods previously discussed, the tasks to be done by candidates are highly structured and standardized as are the rating criteria. These tests are usually based on content validation, and they are usually well accepted by candidates. The cost and complexity of administration of many of these tests, except clerical tests, results in their not being widely used with large candidate groups.

### **Similarities to the Private Sector**

Several other types of tests used by public sector employers are similar to those used in the private sector. Minimum qualification requirements (training, experience, licenses, etc.) are specified for most jobs, sometimes by statute. In the case of public agencies, candidates must meet these requirements in order to receive further consideration. There have been recent attempts to establish the content validity of minimum requirements, but this methodology is still evolving.

Seniority is sometimes used as the basis for selection decisions, but probably less frequently than it is in the private sector. In addition, seniority may contribute additional points to a score based on another type of test rather than constitute the entire test or rank-ordering process.

Assessment centers are used by a number of public agencies, usually as a tool for promotion to high level management or supervisory jobs. Candidates in the public sector are given a numerical score by which they may be rank ordered. Again, this is done to ensure objectivity and fairness.

Performance evaluation is sometimes used as a part of a process to promote employees to higher level jobs. As in other instances, the difference from the public to the private sector is



the standardized, structured way in which it is combined with other measures. It is surprising that despite the intuitive appeal of performance evaluation as a basis for promotion, there is virtually no literature on its validity.

Interviews in the public sector usually form the basis for the final selection of the individual to fill the position from among the highest scoring candidates on the rank-ordered list. In the private sector, a series of one-to-one interviews and resume reviews may be the entire selection process. Selecting officials are often required to justify their selections, so that even the interviews held in the public sector tend to be more structured.

### **T**ypes of Tests Rarely Used in the Public Sector

Personality and interest tests are rarely used as a part of the selection process by public employers. In some cases, these tests are specifically prohibited by law or regulation, such as the prohibition in the Federal Personnel Manual (FPM Chapter 337, subchapter 1 concerning outside hires, and FPM Supplement 335-1, subchapter 3.3 concerning merit promotion). Most commercially available personality and interest tests were not developed for use in personnel selection. These tests are rarely validated for specific jobs. In addition, many of these tests require specialized expertise in administration and interpretation. A high level of expertise and a substantial amount of research is needed to develop these tests. The significant exception is that a number of public agencies use psychosocial screening to select persons for law enforcement positions. The tests are usually administered by a psychologist or psychiatrist to the highest scoring (based on other tests) candidates and are often designed to screen out unacceptable people rather than to select the best. Applicants often appeal decisions based on these tests, so it is usual to employ structured criteria and to use only extreme results as a basis for exclusion.

Biographical data (biodata) inventories are also rarely used by public agencies. (Note that the U.S. Office of Personnel Management is considering use of biodata as part of a test battery for entry level hiring.) Biodata inventories consist of life history questions, often personal in content, which have been demonstrated to relate to job success based on criterion-related validation studies. Often biodata items bear no intuitive or overt

relationship to the job; more direct questions may be subject to faking. Public employers are uncomfortable using such tests because they can be viewed as violating merit principles. Such tests do not meet the expectations of most applicants or managers and can be the subject of strong controversy. In addition, if a jurisdiction must reveal the answer key to its biodata questionnaires (perhaps as a matter of public information), they become all the less acceptable as they become extremely easy to falsify. Also, unlike many written examinations, there are relatively few valid biodata questions, so changes from test to test may be impossible.

Personality, interest and biodata tests all present similar problems for use in public sector personnel selection. Because of their frequently personal nature, questions on these tests often are perceived to be unfair and to constitute an invasion of privacy. As a result, they may be objectionable to some candidates. Continued utility of these questions requires that they not be explained to candidates. This can create serious credibility problems. Unfortunately explainable, obvious items are subject to faking by candidates, which creates validity problems. Additionally, alternate forms of biodata instruments are not available because of the extremely large samples required to develop empirical scoring keys and because of the limited universe of potential biodata items. These problems combine to explain why these tests are rarely used by public sector employers. In a later section of the paper, we call for expanded research on measures such as these to increase their usefulness in the public sector.

---

## ***Areas of Greatest Need for Psychometric Research***

---

There are two topics which stand out from all others as needing reevaluation and progress: measuring the affective domain and selecting supervisory and management personnel. Each is discussed in turn.

### **M**Measurement of the Affective Domain

As discussed earlier in this paper, tests used in the public sector for personnel selection and promotion are generally cognitive in nature. They test knowledge, skills and abilities needed to perform various jobs. Rarely do tests attempt to measure the affective domain — personality or emotions — or even social or political orientation. The earlier discussion of why affective tests such as biodata, personality tests or interest inventories are rarely used for personnel measurement in the public sector indicates the great need for research on innovative techniques in this area of measurement. And, that research must address the special concerns of testing in the public sector.

Validity generalization research has indicated that cognitive tests have validities in the range of .25 to .55 (Schmitt *et al.*, 1984). If we assume that the findings of validity generalization studies are correct, then cognitive tests provide us with valuable information about a candidate's likelihood of success on a job. These studies also indicate that cognitive tests show only part of the picture; a major portion of the factors which lead to job success is not measured by cognitive tests. Day and Silverman's recent (1989) study of accountants supports the view that personality variables are significant predictors of job performance beyond the information provided by measures of cognitive ability.

In discussions with managers and supervisors concerning what makes a successful employee and what factors differentiate the most successful from others, the factors cited include both cognitive and personality characteristics. Personality traits such as reliability, dependability, conscientiousness, ability to get along with others, ability to work as part of a team, interest in the work, self-motivation and willingness to work extra hours when necessary are often cited by managers when discussing job success in a wide variety of jobs.

There is another reason why tests measuring the affective domain are desirable. Cognitive tests tend to have an adverse impact against minority group members. Tests measuring the affective domain, to the extent that there is sound research on this, seem to have little adverse impact (Baehr, 1976, Hogan *et al.*, 1985, and Hough, 1988).

As they are used by the public sector, oral boards, assessment centers, biodata and ratings of training and experience all assess aspects of the affective domain, but their primary emphasis is cognitive. Few personality or interest tests have been designed for use in personnel selection. Rather, they have been developed for research purposes, clinical diagnosis or for use in student or adult counseling. Like diagnostic tools, personality tests and interest inventories designed specifically for use in personnel selection must recognize that the people taking them may wish to present an ideal image they believe will get them a job, not a true picture of themselves. There have been a few such measures which have been developed recently, such as the Hilson Personnel Profile (Inwald, 1988), Vocational Interest Questionnaire (Davey, 1982), and Hogan Personality Inventory (Hogan, 1985). Further research is necessary to establish the validity, reliability, usefulness, and practicality of these measures.

Measurement of the affective domain in the public sector is particularly difficult because those tests appear to applicants to be scored subjectively, without a clear right or wrong answer, and to be easily faked. For non-cognitive tests to be usable in the public sector, these concerns need to be addressed. Creative new measurement ideas are needed to fill an important void. If developed and thoroughly researched, measures of the affective domain could address some of the most significant areas of concern for public personnel assessment, namely an increase in validity and a reduction in adverse impact on minority group members. This is a difficult but necessary undertaking.

## **Selecting Supervisory and Management Personnel**

Many jurisdictions use civil service examinations to help decide whom to promote to supervisory and management positions. Tests used for this purpose may include written tests of supervisory and subject matter knowledges, oral examinations, evaluations of training and experience, or practical exercises and assessment centers of various composition. Despite over 50 years of research and development of personnel selection methods and validation of personnel selection tests and despite the large number of jurisdictions which hold promotional examinations, there is a mere handful of criterion-related validity studies of tests to select supervisory and management personnel in the public sector. There is a similar lack of definitive research in the private sector.

The public sector needs a method of selecting and promoting supervisory and managerial employees that is practical, fair, reliable and demonstrably valid on its face, and based on criterion-related as well as content validity studies. Current practice involves assessing diverse areas using diverse methods. Job knowledge tests are often criticized as not covering all aspects of the job, ignoring the application of the knowledges measured. Traditional training and experience evaluations are criticized for giving undue credit to education and not differentiating among levels or quality of past job performance. Assessment centers resemble an art rather than a science. Methods abound; criterion-related validity studies do not. This section briefly explores the nature of the dilemma of selection of supervisory and managerial personnel and some of the reasons for its still being a dilemma, and briefly critiques one popular method of selection of such personnel.

### **Unique Factors in Selection for Supervisory and Managerial Jobs**

There are at least two reasons why promotion or selection to supervisory and managerial jobs is different from selection to other jobs. First is the change in the nature of the skills needed and the related difficulty in not promoting the "best worker." In a typical promotional setting, say from a Junior to a Senior Computer Programmer, the person with the best programming skills is the logical choice for promotion. But for promotion to a

supervisory position, it may be that a person other than the best programmer should be chosen. This may not sit well with the Junior Programmers who are working diligently with the expectation that good work will lead to promotion. This is one source of difficulty. Second, and more important, is the difficulty we have in describing and measuring the skills which are needed in the supervisory or managerial job.

### **Definition of supervisory and management skills and abilities**

There is little consensus on the KSAPs which are required to perform supervisory and managerial jobs. (The "P" in KSAP stands for "personal characteristics.") This is surely due, in large part, to the multiple ways such jobs can be successfully approached, indicating that there may be more than one set of KSAPs which qualifies one to perform such jobs. It is also partly due to the role leadership plays in these types of jobs and the lack of agreement in the scholarly literature on theories and methods of leadership. For example, one viewpoint suggests, with some empirical support, that policies, rules and other organizational and individual attributes may serve as substitutes for leadership (Kerr and Jermier, 1978.). Even if there were agreement as to the determinants of success as a leader in one type of organization, such agreement might not hold for other types of organizations. For example, it may well be that civilian police departments are qualitatively different from the military. For example, unlike the military, the entry-level police officer has the greatest latitude in carrying out orders (Hale, 1981, p. 23). It is intuitively likely that KSAPs which underlie success as a leader in a highly technical corporation, in a sales organization and in a police department do not overlap completely.

The classic management literature suggests that there are seven management functions: planning, organizing, staffing, directing, coordinating, reporting and budgeting; and some fourteen or more management principles, such as unity of command and chain of authority (e.g., Hampton, 1986, p. 62). The principles of classical management theory were not derived from empirical evidence, and their validity and value has been questioned (Howell & Dipboye, 1986, pages 24-25).

The leadership literature is constantly advancing. The Ohio State studies (for example, Landy, 1985, pages 429-437) identified two major, independent factors of leadership: consideration and structure. One focused on people as human beings and the other on job tasks. This finding is the basis for many widely distributed training programs in supervision. Recently it has been suggested that the Ohio State research may be valid only for supervisors in manufacturing environments and that it was limited in the types of leader behaviors which it studied.

There are a number of well supported theories which deal with aspects of leadership, such as reinforcement theory (summary in Muchinsky, 1987, pp. 470-475), equity theory (summary in Muchinsky, 1987, pp. 456-463), goal setting (Locke, Shaw, Saari & Latham, 1981), and decision making (summary in Muchinsky, 1987, pp. 520-525).

The focus of much research now has shifted to transformational aspects of leadership, considering how and under what circumstances a leader changes rather than satisfies the motivational characteristics of employees (e.g., Bass, 1989; House, 1989).

Simply, the integration of these theories into a verifiable, comprehensive theory has not yet been accomplished. So the proper approach for a leader to use in any situation is a matter without firm guidance.

There is little agreement on what variables of the leader and the situation affect each other, on how to measure these, or on the directionality of the interaction, despite considerable agreement that there is some interaction (Hampton, 1986, pages 467-472; Howell & Dipboye, 1986, pages 191-2).

The matter is further complicated by the absence of a standard technology for describing differences between supervisory or management job levels and assignments, the differences between organizational structures and organizational leadership styles, and the implications of these differences for personnel selection.

### **Technological Constraints on Validity Studies**

There are many reasons for the scarcity of sound criterion-related validation studies for these jobs. In addition to the lack of sound measures of supervisory and managerial job performance, another reason for the recent dearth of criterion-validity studies

is the demonstration, by Schmidt, Hunter and Urry (1976) that large sample sizes are needed for statistically powerful validity studies. Many researchers now hesitate to undertake criterion-related validation studies with fewer than 200 hires. This greatly limits the research base for supervisory and management jobs. Perhaps for this research topic it would be better to have many studies with small sample sizes than no studies at all.

### **Public Sector Research Reports**

There are few reports of criterion-related validity studies for the selection of public sector supervisors or managers in the general (journal or textbook) literature (c.f., Cascio, 1982). Despite their prevalence and importance, only two such studies which deal with police manager assessment centers (McEvoy & Beatty, 1989; Ross, 1980, cited in McGinnis, 1987, p. 109) are known to the authors. Rather than being found in the general literature, criterionrelated validation studies may be more commonly published as technical reports. One such study was conducted by the United States Office of Personnel Management (Corts, 1980) which reported the development of a generic test for the ranking of applicants for trades and labor supervisory jobs. The test covers 31 areas such as:

1. Interest and ability in applying up-to-date job practices
2. Learning and reasoning
3. Flexibility
4. Knowledge of the job as required for a supervisor
5. Checking on work progres
6. Getting information from employees and acting on it
7. Helping employees with personal problems

The uncorrected concurrent validity coefficients ranged from .14 to .60, depending upon the precise criterion considered, based on data from 272 supervisors. Of equal interest for the sake of the present review, that study did not find any earlier studies of a similar nature. In fact, its entire reference section contained only seven items, none relating to similar work by other groups.

McCann Associates has completed two criterion-related validity studies for multiple job knowledge promotional tests for the police and fire services (McCann, Zupkis, Howeth, Nichols, 1975; and McCann, Howeth & Nichols, 1983). They found cross



validated validity coefficients of .25 to .32 and uncorrected validity coefficients as high as .50.

Additional studies which have been published as technical reports are difficult to uncover as there is no systematic method for searching the technical report literature. Informal search methods, mainly word of mouth, indicate that there are few, if any, other such criterion-related validity studies.

### **Approaches to Selecting Supervisors and Managers**

There is a plethora of approaches to selecting people for supervisory and management jobs. In his review of the literature, Cascio (1982) summarizes validity evidence for measures such as: cognitive ability tests (e.g., verbal reasoning and mathematical ability), objective personality and interest measures, projective techniques, biodata, peer assessment and business games, among others. Two of the more popular approaches, assessment centers and employee performance evaluations, will be considered here.

**Assessment Centers.** This method is perhaps most often discussed with respect to supervisory and managerial selection. There are several reasons why the assessment center method is not yet a complete solution to the dilemma of selecting supervisors and managers.

First, the assessment center method includes measurement techniques which are too diverse to be validated or otherwise treated as one selection instrument, perhaps not even as one method. The variation in assessment centers is at least as great as the variation in multiple choice tests, for reasons which go beyond the inclusion of multiple choice tests in some, but not all assessment centers. Assessment center exercises range from measures of knowledge, to measures of personality, to simulations of job tasks. For example, the classic AT&T assessment center (summarized in Cascio, 1982, pages 243-244) lasted three and one half days, included at least nine types of measures (paper and pencil tests, in-basket tests, projective personality tests, clinical interviews with psychologists, group problems, leaderless group discussions, a personal history questionnaire, an autobiographical essay and a self description essay) and rated some two dozen areas (e.g., organization and planning, decision making, creativity, human relations skills, personal impact,

behavioral flexibility, tolerance of uncertainty, resistance to stress, scholastic aptitude and range of interest). In comparison, an assessment center for promotion to police sergeant or lieutenant may last a few hours and consist of an oral presentation, a leaderless group discussion and an in-basket exercise. It is clear that this is not equivalent to the AT&T assessment center. And, due to these differences, experts in the field suggest that public sector (and especially police and fire) assessment centers should be supported by their own criterion-related validation (McGinnis, 1987, p. 108). However, these assessment centers have not been so validated, often due to the technological constraint imposed by small sample size.

Second, there is very considerable variation in what dimensions or areas are targeted for measurement by a given assessment center. These may range in number from a half dozen to a dozen or more areas. The areas measured also vary widely in name and definition and do not seem defensible as psychological constructs. This is particularly troublesome if we must meet the standards for construct validation in the *Guidelines*.

Third, to the extent that personality measures of various types, and projective personality tests in particular, are included in an assessment center, the practical nature of the civil service examination mandated for some jurisdictions is compromised. Further, the acceptance by applicants may be problematic; answering an appeal by telling an applicant that on the Thematic Apperception Test he/she told a story about failure which was not given as much credit as another person's story about success is likely to bring incredulity rather than acceptance.

Fourth, it is not clear what assessment centers measure. Originally they attempted to measure the requisite KSAPs. However, over the past few years it has been noted that the ratings of different areas within a given assessment exercise are more highly correlated than the several measures of one KSAP derived from different exercises (McGinnis, 1987, p. 108, and Bycio, Alvares and Hahn, 1987). Thus, the assessment center measurements may well be exercise specific, rather than reflective of underlying KSAPs.

Fifth, some researchers have suggested that assessment centers do not measure the ability to perform on the job so much as capture the organization's policy for promotion (Klimoski &

Strickland, 1977; Turnage & Muchinsky, 1984). A derivative criticism is that in capturing this policy, the assessment center may be unfair in its ratings.

Sixth, there is a considerable debate as to whether a final overall consensus is better than statistical combination of scores from individual exercises. Despite logical arguments to the contrary, there is evidence that statistical or mechanical composite scores are as effective or more effective than consensual or clinical composites (McEvoy and Beatty, 1989, p. 42). Cascio (1982, p. 250) in his review of the literature cites Sawyer (1966) and others, as showing that statistical prediction is always equal to or better than judgmental prediction. In his review of the literature, Landy (1985, p. 96) reaches the same conclusion.

Seventh, the cost of assessment centers quickly becomes prohibitive in situations with many applicants.

At worst, assessment center methodology is expensive and bewildering in its complexities, and the results are suspect. At best it is the most fair and valid approach to supervisory and management selection. Unfortunately, with the current level of scientific knowledge and with the current technology, we cannot say where the assessment center method falls between these two extremes.

Employee Performance Evaluation. For many people, common sense would suggest that we base promotion decisions on past job performance. There are many reasons why this is not a practical approach, beyond the difficulty mentioned above concerning the match between current job duties and the partially unique job requirements of supervisory and managerial jobs. Objective measures of job performance are difficult to find for many public sector jobs. Even worse, putting emphasis on the quantitative aspects of job performance, as occurs in most Management by Objectives or MBO systems, detracts from the qualitative aspects of job performance (Hampton, 1986, page 155). For reasons such as these, performance evaluation usually is based on the subjective judgment of supervisors. Unfortunately, subjective evaluations are open to various types of unintentional errors (such as leniency, halo and central tendency) as well as personal biases and intentional manipulation. Further, applicants for a promotion may be serving in different job titles, and comparing the job performance of people in different job titles is

difficult. For example, consider two applicants for a managerial position as Director of Computer Systems. The first is a project supervisor for development of main-frame computer systems. The second is in charge of the department's personal computers, including purchasing of hardware and software, training the department's employees and maintenance of the personal computer system. It is not clear whether "good" performance on the part of these two applicants is comparable or whether one or the other applicant has the harder job. Finally, performance evaluation systems are often thought unfair and strongly resisted by employee and supervisor alike. For these reasons and others, employee performance evaluation is rarely a tenable approach to selecting supervisory and managerial personnel in a merit system.

### **Suggestions for Further Research**

As mentioned earlier, a major problem in conducting criterion-related validation studies for supervisory and managerial jobs is the typically small sample size. While this has led to avoiding such research, it is suggested that it should now lead to innovative approaches to the problem. For example, it may be possible to conduct the requisite research with smaller sample sizes than previously thought necessary if the best and the worst applicants were hired. This is not as outrageous as it may initially appear, since any criterion-related validation study strives to include data from all points along the test and job performance continuum.

The criterion problem may be addressed, in part, by a new nationwide impetus for performance evaluation at all levels in the police departments. An accreditation program being conducted by the Commission on Accreditation of Law Enforcement Agencies in Fairfax, Virginia, is encouraging performance evaluation and validated selection procedures in police departments across the nation. As a result, many police departments will become familiar with and begin using systematic evaluation of their personnel at all ranks. With the obvious benefits of this possible consistency comes the need to find a way to assess the effects of departmental differences in organizational structure, culture, mission, and social, legal and political context.

## **Need for a Technology of Managerial Selection**

The public sector has an immediate need for an agreed-upon technology of selection for supervisory and management personnel. The method must meet the usual psychometric/social/legal requirements of reliability, validity, utility and fairness, and ease of use. Of course, it must reflect public policy as determined by the lawmaking and policy making bodies of the country, such as lowest possible adverse impact, defensibility under the *Guidelines* and related state and federal equal employment opportunity law, rules and regulations. It must be capable of secure use, offering no applicant an unfair advantage, either in reality or in perception. It must also be intuitively valid to applicants and other interested parties (e.g., news media), and capable of objective scoring. Finally it must meet the requirement of many civil service jurisdictions that the test be practical in character and deal in so far as possible with the actual duties of the position.

The existing literature in this area does not meet this ideal for two reasons: leadership and supervision are not fully understood constructs, or sets of constructs, and the field has not yet agreed upon common definitions for the germane areas to measure nor which tests or types of tests to best measure them. Despite this lack of scientific clarity, personnel assessment professionals in merit systems are continually being asked and even required to develop sound and defensible procedures for promoting or selecting people for supervisory positions and to do so without unnecessary expense. This serves to emphasize the need for integration of the literature and publication of practical examples of assessment instruments for promotion or selection to supervisory positions. An excellent model for the second part of this is the text edited by Gael (1988) which gives both theoretical treatment and practical examples of job analysis techniques.

The integration of the literature should result in a virtual technology or guidebook to accepted practice which deals with such topics as:

- definition of assessment areas, reasons for differences in definitions, and the degree of and the need (if any) for factorial purity of such areas.
- relative weight to be given job knowledge, general cognitive ability, supervisory/leadership skills, interpersonal skills and personality variables.

- whether and how the weights given to the various components should vary with the nature of the job.
- which job requirements can be reasonably expected to be learned on the job, such as job knowledge, general cognitive ability, supervisory/leadership skills, interpersonal skills, personality variables.
- relative merits of situation-specific test questions as compared to general or pure tests of cognitive ability or problem solving.
- relative merits of tests of knowledge and understanding of principles and practices of supervision and leadership versus other approaches to measuring these areas, such as written or video presentation of situational questions or simulation exercises.
- the appropriate deference to be given to the management style of the organization. That is, should a selection process for an autocratic organization be different from that for a participative organization?
- the extent to which grading criteria determined by groups of subject matter experts (SMEs) are consistent over time, consistent across groups of SMEs, and objectively correct.

There is a strong need for a synthesis of the scientific literature concerning selection of supervisory and management personnel and the development of an agreed-upon technology for using that body of scientific knowledge as it now exists. Other disciplines with fast changing bodies of knowledge (e.g., medicine) develop technologies for practice. Whatever the level of scientific knowledge, public sector personnel selection needs to conduct personnel selection today and must apply that body of knowledge. Nothing less than an accepted technology of managerial testing is needed, based on sound psychometric and psychological theory and research. Coordinated advances are needed in the technology of management testing and the scientific theory of organizations in the area of management selection.

---

## ***Call for Systematic Funding of Research and Development***

---

This paper has presented the unique constraints on personnel testing in the public sector. With much of the structure placed on public employers by laws and regulations, innovation, research and change are often difficult and slow. But, despite the many constraints, there have been numerous innovations. The IPMAAC Monograph *Recent Innovations in Public Sector Assessment* by Charles F. Sproule outlines many of these.

### **Need for Legislation Encouraging Research**

There have been several legislative efforts to encourage research and innovation. In the Federal Civil Service Reform Act of 1978, a provision was made for research and demonstration programs. The law allowed up to ten demonstration projects to be carried on at a time. Such legislation encourages research and innovation. While this legislation covers the entire field of personnel management, personnel testing is clearly included. At the state level, 1987 legislation in Minnesota allowed waiver of personnel rules or statutes for experimental or research projects designed to improve personnel recruitment, selection, referral or appointment procedures. While there are significant limitations on the scope of the research, this legislation allows for research free of many of the normal constraints. Such research enables public agencies to prepare sound bases for presenting revisions to laws and regulations.

Laws such as these encourage innovation in personnel testing. With enabling legislation, public employers may experiment more freely, working on research similar to that conducted in the private sector. The authors urge the Commission on Testing and Public Policy to consider adopting a position favoring legislation encouraging and enabling creative innovative research in personnel testing by public sector agencies.

## Need for Research Funding

In 1970 Congress passed the Intergovernmental Personnel Act (IPA). The law authorized federal grants to state and local governments to carry on innovative projects. Under IPA funding, the research components of many state and local personnel testing agencies were begun or grew significantly. Much innovative work in personnel testing was done in state and local governments from 1970 to 1981. The IPA grant program had been intended to be funded at the level of \$20 million for each year from fiscal year 1972 to fiscal year 1986 (Congressional Record, 1981). However, these funds were completely eliminated in 1981. With the elimination of IPA funding, these functions have been severely cut back.

With support from IPA, consortia of state and local personnel testing agencies were founded. By the later 1970s and early 1980s, there were consortia in almost every part of the country: New England Public Personnel Council (NEPPC); Mid-Atlantic Personnel Assessment Consortium (MAPAC); South-East Region Personnel Assessment Consortium (SERPAC); Great Lakes Assessment Council (GLAC); Missouri, Iowa, Nebraska, Kansas Selection Information Exchange (MINKSIE); Region Six Personnel Assessment Consortium (RESPAC); and Western Region Personnel Assessment Consortium (WRIPAC). These groups met periodically to exchange information and work on cooperative projects in such areas as staff training, validation and item banking. In 1989 only three of the consortia, WRIPAC, MAPAC and GLAC, continue to exist.

IPMAAC began in 1976 as a result of a view that the consortia at the regional level were successful and that an organization was needed to perform a similar function on a national level. While IPMAAC never received direct IPA funding, many of its members were involved in projects supported by IPA.

Many noteworthy projects were done by state and local jurisdictions in the area of personnel testing with IPA funds. Much of the initial research was conducted on improving tests for use in selection and promotion in the Police and Fire service. A *Content Validity Manual* was prepared by the City of Minneapolis (Musio and Smith, 1972) which is still used today. Initial efforts at item banking, exploration of alternatives to traditional written



tests, development of multi-purpose job analysis, and research on many other areas were begun under IPA funding.

In addition, under IPA, the U.S. Civil Service Commission, which was later renamed the U.S. Office of Personnel Management, provided a substantial amount of technical assistance to state and local agencies. Besides the Personnel Research and Development Center in Washington, with a large staff of industrial psychologists, each of the regional offices contained a staff of one or more psychologists who assisted federal, state and local agencies on personnel testing issues. Virtually all of the Regional Psychologists now have been eliminated and the Personnel Research and Development Center has been very severely cut back in staff. The U.S. Office of Personnel Management now provides little assistance to state and local agencies.

As late as 1979, the U.S. Office of Personnel Management was responsible for the monitoring of compliance of state and local programs under detailed Federal Merit System Standards (U.S. OPM., 1979). The 1979 standards provided more detail than the current (U.S. OPM., 1983) merit system standards and incorporated as a requirement the stringent *Uniform Guidelines on Employee Selection Procedures*. In 1983 the U.S. Office of Personnel Management changed its mode of monitoring compliance. Now a notice from the chief executive of the jurisdiction is taken as indicating a jurisdiction's compliance. In some jurisdictions this change directly resulted in a withdrawal or reduction of support for merit system programs.

The authors urge the Commission on Testing and Public Policy to consider calling for a restoration of funding for the Intergovernmental Personnel Act to provide the resources needed to continue advancement in personnel testing in the public sector.

---

## *Conclusion*

---

At one point public jobs were bought and sold in pubs and in the halls of legislatures and Congress. Largely as a result of public outcry at various times over the past century, many government jurisdictions now use formal merit systems to select and promote their employees. These merit systems have many checks, balances and constraints codified in law, rule, policy and procedure. They are the epitome of a bureaucracy: run in accordance with numerous and relatively inflexible rules but treating all concerned openly, equally and fairly, albeit slowly. Unfortunately, the same rules which promote equal treatment have also become entrenched and difficult to change. And while there is extremely limited funding to promote change, there is a very complex professional and legal literature concerning employee assessment and selection, some of which suggests that change is warranted.

### **S**ummary

According to its statement of purpose, the National Commission on Testing and Public Policy is “conducting a three-year study of the role of testing in the allocation of educational, training, and employment opportunities. The major outcome ... will focus on the social, economic and political contexts, uses, and consequences of testing.”

This paper contributed to this purpose by presenting a public sector viewpoint on assessment issues. We discussed:

- Legal mandates and public expectations of the public sector, including the effect of merit principles
- Common characteristics of federal, state and local government
- Differences in levels of government
- A comparison of selection in the public and private sectors

We also presented special conflicts in public sector testing, including conflicting statutory mandates, conflicting statutory and political mandates, conflicting legal and professional standards, pressures from applicant groups, conflicting loyalties of

human resource professionals, and factors contributing to such conflicts.

A third area of discussion concerned constraints on the use of testing methods and included testing techniques, similarities in methods to those used in the private sector, and types of tests rarely used in the public sector.

A fourth section focused on the areas of greatest need for psychometric research: testing the affective domain and designing tests to predict supervisory and managerial job performance.

Finally, the need for legislation to encourage innovative research and for systematic funding of testing research was presented.

## **Recommendations**

The last two sections of the paper focused on areas of need for research and support for that research. To allow these government merit systems to improve involves overcoming these four common difficulties:

- poor and uncertain funding for research
- inadequate base of knowledge and technology, particularly regarding testing the affective domain and predicting supervisory and managerial job performance
- inadequate training of assessment staff
- laws which constrain change

All of these can be remedied. The first three were improving rapidly until the federal government eliminated the IPA program. This program funded research on methods of testing, development of new and improved test development methods, and training of assessment staff. If Congress passes a new IPA program progress in personnel selection at the state and local level will be dramatic. Much of the foundation for such change is in place. A new IPA program would provide the mechanism for refining and disseminating these improvements. A fresh look at the many statutes concerning civil service hiring is impossible without a specific mandate to do so and a model of how to proceed. The modern civil service agency walks a tightrope between the right of the community to have the best government work force and the right of the individual to be fairly considered for employment. Legislation is needed which will allow the merit system to

continue while fostering maximum openness and fairness in the civil service hiring process. So, to address the last difficulty area, model legislation should be developed by an influential body to demonstrate a reasoned and legitimate degree of flexibility in a merit selection program. This model legislation must then be advocated by federal and state agencies.

---

## *References*

---

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Anastasi, A. (1988). *Psychological Testing*, 6th ed. New York: Macmillan Publishing Company.
- Ash, R.A. & Levine, E.L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology*, 70, 572-576.
- Ashton, D.P. & Wiesen, J.P. (1983, May). *Minimizing adverse impact while maintaining a merit system*. A paper presented at the International Personnel Management Association Assessment Council Annual Conference, Washington, D.C.
- Bass, B. (1989, April). *The current state of leadership theories: Integration, research, applications*. Panel discussion conducted at the meeting of the Society for Industrial and Organizational Psychology, Boston.
- Baehr, M.E. (1976). *National Validation of a Selection Test Battery for Male Transit Bus Operators*. Washington, D.C.: US Department of Commerce, National Technical Information Service Report Number UMTA-MA-06-0011-77-1.
- Bycio, P., Alvares, K.M. & Hahn, J. (1987) Situational specificity in assessment center ratings: a confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463-474.
- Campion, M.A., Pursell, E.D. & Brown, B.K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology*, 41, 25-42.
- Cascio, W. F. (1982). *Applied Psychology In Personnel Management*, 2nd ed. Reston, Virginia: Reston Publishing Company.
- Congressional Record (April 29, 1981) Legislation to abolish the Intergovernmental Personnel Act, p. S 4140, column 2, paragraph 4.
- Corts, D. (1980, October). *Development And Validation Of A Test For The Ranking Of Applicants For Promotion To First-Line Federal Trades And Labor Supervisory Positions*. Personnel

- Research Report 80-30, Washington, D.C.: U. S. Office of Personnel Management.
- Connecticut v. Teal, 457 U.S. 440 (1982).
- Day, D.V. & Silverman, S.B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, 42, 25-36.
- Davey, B. (1982). *Development and validation of an interest inventory for police selection*. A paper presented at the International Personnel Management Association Assessment Council Annual Conference, Minneapolis, Minnesota.
- Executive Order 11246, 30 Fed. Reg. 12319 (1965) as amended by Executive Orders 11375 (1967), 11478 (1969), and 12086, 43 Fed. Reg. 46501 (1978).
- Federal Personnel Manual (current edition). Washington, D.C.: U.S. Office of Personnel Management.
- Flaughner, R.L., Nieves, L., Slaughter, C., Wiesen, J.P. & Woodford, P. (1980). *An Approach For Identifying and Minimizing Bias In Standardized Tests: A Set of Guidelines*, Office for Minority Education Monograph Number Four. Princeton: Educational Testing Service. (Originally published in 1979 by the Massachusetts Department of Personnel Administration, Boston.)
- Gael, S. (Ed.) (1988). *The Job Analysis Handbook for Business, Industry and Government*, Volumes I and II. New York: Wiley.
- Griggs v. Duke Power Co. 401 U.S. 424 (1971).
- Hale, C.D. (1981). *Police Patrol: Operations and Management*, New York: John Wiley.
- Hampton, D.R. (1986). *Management* (3rd ed.) New York: McGraw-Hill.
- Hartigan, J. A. & Wigdor, A. K. (Eds.) (1989) *Fairness in Employment Testing*. Washington, D.C.: National Research Council, National Academy Press.
- Hogan, R. (1985). *Psychological screening for bomb technicians*. A paper presented at the International Personnel Management Association Assessment Council Annual Conference, New Orleans, Louisiana.
- Hogan, R., Carpenter, B.N., Briggs, S.R. & Hanson, R.O. (1985). Personality assessment and personnel selection. In *Personality*

- Assessment in Organizations*, H. Bernardin and D. Bownas (Eds.) New York: Praeger.
- Hough, L. (1988). *Personality Assessment for Selection and Placement Decisions*. Minneapolis, Minnesota: Personnel Decisions Research Institute.
- House, R. (1989, April). *The current state of leadership theories: Integration, research, applications*. Panel discussion conducted at the meeting of the Society for Industrial and Organizational Psychology, Boston.
- Howell, W. & Dipboye, R. (1986). *Essentials of industrial and organizational psychology* (3rd. ed.) Chicago: Dorsey Press.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J.E., Schmidt, F.L. & Rauschenberger, J. (1984). Methodological and statistical issues in the study of bias in mental testing. In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum.
- Inwald, R. (1988) "Screen-out" v. "screen-in": *Two models for pre-employment psychological testing*. A paper presented at the International Personnel Management Association Assessment Council Annual Conference, Las Vegas, Nevada.
- Janowiak v. Corporate City of South Bend, 836 F.2d 1034 (7th Cir. 1987).
- Johnson v. Transportation Agency of Santa Clara Co., 480 U.S. \_\_\_, 108 S. Ct. \_\_\_ 94 Law Ed. 2d 615 (1987).
- Kerr, S. & Jermier, J.M. (1978). Substitutes for Leadership: Their meaning and measurement. *Organizational Behavior and Human Performance*, 22, 375-403.
- Klimoski, R.J. & Strickland, W.J. (1977). Assessment centers-valid or merely prescient. *Personnel Psychology*, 30, 353-360.
- Landy, F.J. (1985). *Psychology of Work Behavior*. Homewood, Illinois: Dorsey Press.
- Locke, E.A., Shaw, K.N., Saari, L.M. & Latham, G.P. (1981). Goal Setting and Task Performance: 1969-1980. *Psychological Bulletin*, 90, 125-152.

- the assessment paradigm. *Journal of Applied Psychology*, 69, 595-602.
- Uniform Guidelines on Employee Selection Procedures*. (1978), 29 C.F.R. Part 1607. (Also: 44 FR 11996, March 2, 1979 and 45 FR 29530, May 2, 1980.)
- U.S. OPM. *Standards for a merit system of personnel administration: Final Rule; Revision*. *Federal Register*, Friday, February 16, 1979, part III, 10238-10264.
- U.S. OPM. (1983). *Intergovernmental Personnel Act Programs; Standards for a merit system of personnel administration, 1983*. 5 C.F.R. Part 900.
- Wards Cove Packing Co. v. Atonio, U.S. Sup. Ct., No. 87-1387, 6/5/89.
- Watson v Republic National Bank of Fort Worth. 101 L.Ed. 2d 827, 108 S.Ct. 2777 (1988).
- Wiesen, J.P. (1988). *The Connecticut Merit Promotion System: A Program Evaluation And Some Recommended Improvements*. Hartford, Connecticut: The Personnel Division of the Connecticut Department of Administrative Services.
- Wiesen, J.P., Reynolds, J.J. & Estes, S. (1979) Cultural bias review of test content: quantitative and qualitative: a police sergeant examination. *Journal Supplement Abstract Service*, 9, ms. 1919.
- Wright, G.H. (Ed.) (1974). *Public sector employment selection: A manual for the personnel generalist*. Washington, D.C.: International Personnel Management Association.
- Wygant v. Jackson Board of Education. 476 U.S. 267 (1986). 42 U.S.C. Section 1983.