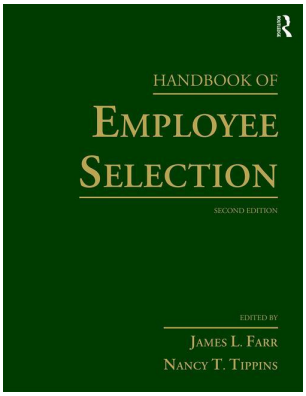
This article was downloaded by: *10.2.97.136*
On: 26 Sep 2023
Access details: *subscription number*
Publisher: *Routledge*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Employee Selection

James L. Farr, Nancy T. Tippins, Walter C. Borman, David Chan, Michael D. Coovert, Rick Jacobs, P. Richard Jeanneret, Jerard F. Kehoe, Filip Lievens, S. Morton McPhail, Kevin R. Murphy, Robert E. Ployhart, Elaine D. Pulakos, Douglas H. Reynolds, Ann Marie Ryan, Neal Schmitt, Benjamin Schneider

Public Sector Employment

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-32
Rick Jacobs, Donna L. Denning
Published online on: 22 Mar 2017

How to cite :- Rick Jacobs, Donna L. Denning. 22 Mar 2017, *Public Sector Employment from:* Handbook of Employee Selection Routledge
Accessed on: 26 Sep 2023
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-32

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://test.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

32

Public Sector Employment

RICK JACOBS AND DONNA L. DENNING

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

Historians often cite the origin of civil service or public sector testing as far back as 2200 BC, when a Chinese emperor used a process of systematic assessment to determine if his officials were fit for office (DuBois, 1970; Frank, 1963). In these early times, individuals were assessed with what might now be labeled job-relevant work samples; they included tests of specific skills such as horsemanship and archery. The Han Dynasty (202 BCE to 200 CE) is credited with moving testing from the actual actions required on the job to a surrogate, written format that included five areas of knowledge: civil law, military affairs, agriculture, revenue, and geography (Gregory, 1996). Candidates who were successful in meeting rigorous cutoff scores on local examinations were deemed appropriate to continue with the process of testing at regional and higher levels in the overall process. Thus, in many respects, these ancient tests were prototypes of what has become known generically as civil service examinations or, more generally, public sector testing and can be seen as way to guard against the potential negative consequences of patronage as well as embrace the positive results of having standardization and more accurate indicators of future performance.

This brief historical description depicts the genesis of public sector testing and illustrates that it shares some similarities with current practices, but important differences exist. Most noteworthy, use of these early tests was deficient in terms of systematic evaluation of outcomes: demonstration of their predictive validity. Furthermore, the tests were conducted under extreme conditions that required candidates to spend long hours in confined spaces that would never be tolerated today, and they routinely had failure rates that were considerably higher than would often prove viable today, well in excess of 90%.

Moving forward 2,000 years, from China (AD 200) to France (1791), England (1833), and finally the United States (1883), we see the more immediate historical roots of modern-day public sector testing (Graham & Lily, 1984). In these systems, tests were used to select individuals for government positions in a way that was intended to be free of patronage and fair to all candidates. These tests were each designed to identify the individuals who were most likely to succeed in a given position on the basis of specific subject matter that made up the content of the tests, a precursor to what is now routinely labeled as validity based on test content. Although much has been done over the years to improve the characteristics of these assessments, such as more carefully matching test materials to job requirements, further standardizing testing processes, and evaluating predictive efficiencies by validation studies, the basic ideas underlying civil service examining have a long and rich history, in fact, one that long predates emergence of the discipline of industrial psychology.

This chapter provides details of the distinctive characteristics of testing in the public sector. It starts with the process of identifying the positions that are part of a competitive examination process and then moves on to discuss the development and administration of entry-level

examinations. The following section reviews validity, or linking tests to jobs. The next section addresses recruitment of candidates; optimal selection decisions require maximizing the number of individuals competing for the job. Part five of this chapter turns to testing for promotional opportunities. Next is a discussion of legal considerations surrounding testing in the public sector. The chapter concludes with a summary of how public sector testing has evolved through the past century and a view on where it might be evolving to in the 21st century.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

POSITION CLASSIFICATION IN THE PUBLIC SECTOR

To fully appreciate the extent to which use of formal civil service examinations is entrenched in public sector employee selection, the role of position classification in the public sector must be considered. In this context, a "position" is the segment of work to be performed by one person. Classification of positions involves documentation and analysis of the work of each position, then grouping the positions with sufficiently similar work into a "class" of positions. More formally, a "class" may be defined as follows:

> a group of positions . . . sufficiently similar in respect to the duties, responsibilities, and authority thereof that the same descriptive title may be used with clarity to designate each position allocated to the class, *that the same requirements as to education, experience, capacity, knowledge, proficiency, ability, and other qualifications should be required of the incumbents, that the same tests of fitness may be used to choose qualified employees*, and that the same schedule of compensation may be used.
>
> (Committee on Position-Classification and Pay Plans in the Public Service, 1941, p. 45, italics added)

Historically, this has been a judgmental exercise, but more recently it may include use of formal job analytic methods. Subsequent to the creation of a class, additional positions are "allocated" (assigned) to the class when an added need for comparable work is determined and documented and the additional work is deemed sufficiently similar to the work performed by incumbents in the class to warrant inclusion of the position into the existing class. Similarly, an existing class is abolished when the need for the work performed by those in the class no longer exists.

A description of the work performed by incumbents in a class and the qualifications necessary for performing this work are then documented in a "class specification." The "position classification plan" of the organization, then, "consists of (1) the system of classes and class specifications and (2) a code of formal fundamental rules for installation and maintenance of the classification plan" (Committee on Position-Classification and Pay Plans in the Public Service, 1941, p. 47).

The classification of positions is a formal process that provides the underlying rationale for assessment. With the specification of position qualifications and requirements, organizations can seek to identify existing tests or construct new tests that match this information. What we have in this approach are the roots of a content-based validation strategy, which may then stand on its own or may be supplemented by additional validation information such as criterion-related evidence of validity.

CIVIL SERVICE EXAMINATIONS

The provision in the definition of "class" that all positions in it require comparable qualification led to regulatory provisions regarding the evaluation of qualification. The U.S. Code (Section 2301, Title 5), which governs the U.S. federal civil service system and serves as a model for many other government agencies, stipulates that "selection and advancement should be determined solely on the basis of relative ability, knowledge and skills after . . . competition" (i.e., competitive examination). The Charter of the City of Los Angeles is even more explicit on this point, stating that "Examinations shall . . . test the relative capacity of the persons examined to discharge the duties of the class" (The City of Los Angeles, 2009, p. 69).

A separate civil service examination (either a single test or often a series of tests, the scores on which are combined in a specific way to form a final examination score) is typically conducted for selection into each class. Results of the examination appear as a list of candidates who successfully completed all portions of the examination, ranked in descending order by their score. This list is variously referred to as an "eligible list" or "register of eligibles," indicative that all persons on it are eligible for employment in the class on the basis of their having demonstrated in the civil service examination appropriate qualification to occupy a position in the class.

Adoption of the eligible list/register of eligibles, usually by a civil service commission or the head of the department responsible for examining, marks the end point of the examination; but the *selection process* has not concluded, because no one has yet been hired or promoted. This final step is accomplished by the department with the vacancy requesting a "certification" of the list. Then, in accordance with specific, strict rules, a designated number of candidates' names are provided to the department for their final hiring consideration (technically, they are provided to the "appointing authority," who is typically the department head and is the only person who can legally fill a position in a civil service class). This certification rule has many variants and can range, for example, from a "rule of one" (the highest scorer only, who is then hired unless there is reason not to do so, in which case the second highest scorer is considered, and so forth), a "rule of three" (the three highest scorers), a "rule of $2N + 1$" (2 times the number of vacancies, plus 1, of the highest scores), to the very liberal "rule of the list" (all persons on the list may receive final consideration for selection). Evaluation of candidates for this final selection decision is to be based on additional job-related criteria not found in the testing process. These can be difficult to identify when a thorough examination has been given, one that includes not only specific job-based knowledge but also other components such as work-based behavioral measures, personality indicators, and/or biographical information such as work history.

In 1983, voters in Los Angeles approved a City Charter amendment for use of a rule of "Three Whole Scores" for certification selection. This rule accomplished two things: (1) the rounding of scores to whole numbers eliminated the miniscule, decimal point differences in scores that had previously separated the ranks of candidates, and (2) the hiring department was able to consider an expanded pool of candidates for final selection. In all instances described, all candidates tied at a given score are treated the same (either certified for final hiring consideration or not), so rounding scores to whole numbers has a greater impact than might be expected in grouping candidates at a given score (rank) and thus expanding the pool from which a selection can be made.

Civil service examinations are seen as embodying "merit principles" in selection in that they are based on job-related criteria and provide a ranking of candidates in terms of their relative degree of qualification. Position classification invariably results in a class specification document that at the very least provides a starting point for construction of a job-relevant examination. The description of the job in the class specification document is often supplemented by a more detailed job analysis. Job analysis procedures may vary but have the common objective of specifying the work performed (tasks) on the job. Additionally, and especially relevant for the purpose of developing selection testing, the job analysis also often provides identification of the knowledge, skills, abilities, and possibly other personal characteristics needed to perform these tasks. This information then allows for designation of the most appropriate types of tests for use and their content.

Once again, these provisions require that examinations are based on job requirements or the ability to "discharge the duties of the class," which logically results in a content-based test construction strategy. This, coupled with classification practices that are based on extreme similarity of the work required of all positions in a class, results in nearly universal reliance on content-based testing. And not incidentally, the narrowness in scope of the work performed by incumbents in a class thus precludes local empirically based test validation strategies due to limited sample size.

Testing for a Multiplicity of Jobs

The statutory requirement that an objective assessment program be in place for each job (class) and ensuing mandates that examinations be tailored to the unique demands of each are major

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

challenges facing public sector organizations; usually these organizations have a very large number of classes relative to the number of employees. As an example, in one county in the state of Ohio, the public library service employs just over 1,000 individuals, and these 1,000 employees are classified into more than 110 civil service classes. The turnover rate in this organization is approximately 6% annually, indicating that in any given year there may be about 60 job openings, and these 60 openings may span nearly as many classes, each requiring a different civil service examination (State of Ohio, personal communication, February 2008). Similarly, in a medium-size city in Pennsylvania, the Civil Service Commission must monitor staffing for more than 400 classes. Although some testing for larger classes is predictable and regular, many jobs may have a single vacancy only occasionally (and unpredictably), and the organization must be ready to examine candidates for every one of them at any point in time. In the City of Los Angeles, the Personnel Department is responsible for testing nearly 1,000 classes. The sheer number of jobs and requisites of the civil service system creates a situation in which tailoring selection programs very specifically to each job can make timely completion of the development of all examinations extremely challenging.

Another factor contributing to the volume of civil service examinations that must be developed is the reluctance within many agencies to reuse tests. Test security reigns supreme, given the high stakes of these examinations and the need to preserve the integrity of the process, to the extent that considerable caution is exercised even with respect to repeat exposure of test material. And this caution is well founded; incidents of candidates colluding to reproduce a test (by each memorizing a specific set of items) have been repeatedly encountered.

Defining Test Content

One approach that helps meet the demand for a separate examination for each job, given the multiplicity of jobs within a given organization, is a systematic approach of analysis across jobs with a focus on the commonality among jobs. This helps organizations bring order to jobs in terms of their similarities and, potentially, to assessment tools and processes. Such commonalities are identified by analyzing individual jobs and then comparing them for patterns of similar tasks, duties, responsibilities, and/or, most importantly, knowledge, skills, abilities, and other characteristics (KSAOs). Public sector organizations that must select for many classes can help reduce the burden of creating a complete examination unique to each class by constructing assessment procedures and processes for use across multiple classes on the basis of their similarities. This not only makes the development of selection systems more efficient, but such a process can also result in the compilation of normative information for a much larger sample of individuals, which, in turn, can improve the understanding of the tests used and the applicants being considered for employment.

Implementation of such a process requires use of job analysis procedures that are consistent across jobs and that yield results that allow for comparison of jobs. Once this is accomplished, tests that meet the needs of multiple jobs may be created and any modifications for individual jobs can be made. This approach can also simultaneously facilitate consideration of a candidate for multiple jobs through administration of a comprehensive battery of tests. From this perspective, either the candidate, through application to multiple positions with similar requirements, or the organization, via evaluation of candidates for multiple positions simultaneously, can benefit from knowing the relationship among jobs. As earlier stated, for many public sector organizations, the number of jobs considered distinct (i.e., classes) is daunting, and the use of common testing across jobs can help make far more attainable the ultimate goal of timely administration of a formal examination for each class with a vacancy (or to always have an eligible list available for each class).

Although many public sector organizations continue to use traditional job description and job analysis procedures to define jobs, the past decade has seen a rise in the use of competency modeling as an underlying process for identifying job requirements, parallel to its use in the private sector. A competency model may be constructed for higher-level jobs, especially those that are considered leadership positions (Hollenbeck, McCall & Silzer, 2006), or for all jobs in the

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

organization. In both cases, the competencies identified form the basis of the examination plan (Rodriguez, Patel, Bright, Gregory, & Gowing, 2002).

LINKING TESTS TO JOBS: TOOLS AND PROCESSES FOR IDENTIFYING STRONG CANDIDATES

Any examination used for employee selection, whether it takes advantage of similarities identified across jobs or not, must have a logical framework demonstrating how the tests are linked to the job or, more generally, an evaluation of test validity. Note that regardless of the validity evidence used to support the tests included in the selection process, the examiner (or examination analyst, as they are often called) must engage in developing an examination plan. Examination plans link the information about the job to the types of assessments that are included in the selection process. Examination plans provide a logical underpinning not only to the use of a given type of test and its content but also to the weight that each test receives in the final examination score. As an example, in police officer selection, there has been a movement to establish a more broad-based assessment consisting of not only cognitive ability but also personality characteristics, and experiential information that lead to effective policing. An examination plan for the class of police officer would likely include multiple types of tests with specific assessment dimensions for each and instructions as to how these tests are to be considered (pass/fail or weighted) in the final composite score on which candidates are ranked. Following this logic, it is not hard to see that very different jobs (e.g., library clerical worker, meter reader, lifeguard, and purchasing agent) would have examination plans that differ from police officer and from one another, given the nature of each job and the KSAOs necessary to perform in the position.

Public sector employment covers a very wide range of jobs and thus requires use of a correspondingly wide range of assessment tools. Although the final examination may differ for various positions, a similar process is used for examination development for the vast array of public sector jobs. First, an analysis of the job is undertaken (details of job analysis methods are in Chapter 6, this volume). Then, based on the results of a job analysis, an examination plan is developed that identifies the optimal (and feasible) type(s) of test(s) necessary to assess the knowledge, skills, and/or abilities and aptitudes critical to performance of the job. For library clerical workers and meter readers, tests might focus on attention to detail, whereas for lifeguards the certification of successful completion of a first aid course might be supplemented with a physical abilities test that includes water rescue.

Minimum Qualifications

Threshold requirements, usually in the form of education, training, experience, or certification attained, are often established as minimal qualifications for potential applicants. Public sector organizations rely heavily on minimum qualifications as an initial step in the employment process. Minimum qualifications (alternatively referred to as "requirements") are threshold requirements that potential applicants must meet to participate in the competitive examination. In reality, they are the first "test" in the examination, because they consist of carefully established job-related criteria that each applicant must meet precisely to be allowed to proceed further in the examination process. These criteria are clearly communicated to applicants (so those lacking can self-select out at the earliest possible time), consistently and rigidly applied, and are often subject to verification. Their use is completely consistent with the content-based approach to testing that is so prevalent in the public sector, in that the criteria individuals must meet to participate in the examination for a given class are those that indicate a reasonable likelihood that they will have acquired the knowledge, skills, and abilities that will be subjected to more refined assessment through the remainder of the examination. Examples of minimum qualifications run the range of different characteristics such as age for air traffic controllers (both minimum

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

and maximums are specified by the Federal Aviation Administration); education at a specified level; a particular type of license, such as a commercial driver's license for a bus driver; and in the case of police officers, in some jurisdictions the absence of a felony conviction.

Identifying Potential Selection Tools

Once this (preliminarily) qualified pool of applicants is established, public sector testing personnel identify or construct selection instruments that can be used for more refined assessment in the remainder of the examination. They may search the Internet, the professional literature, test publisher catalogues, and/or professional volumes that review tests, such as *Tests in Print* and *Mental Measurements Yearbook* (Murphey, Plake, & Spies, 2006; Spies, Plake, & Geisinger, 2007). At times, this search process results in identification of instruments that are appropriate and sufficient in their coverage to constitute the entire examination used for selecting the most qualified applicants. However, even when this is the case, test security issues may dictate that the use of a test that is readily available may be inappropriate because some candidates may gain access to the tests, whereas others cannot. However, in many instances, certain features of the job or the need to address specific issues of job content require the creation of new tests; in fact, this is often a primary responsibility of the public sector testing professional. Clearly, the development of new assessment tools requires a great deal of time, effort, and skill, and when that is multiplied by the number of jobs in the organization, the workload can become overwhelming. Many public sector organizations pursue another option in some cases by outsourcing to individuals or consulting firms specializing in instrument development. This is especially true for high-stakes positions in which many jobs are being filled and the likelihood of follow-up objections and legal action on the part of candidates is high.

Role of the Interview

As in the private sector, interviews are an extremely common type of test used in the public sector. As with other types of tests, public sector organizations most often use interviews that are carefully tailored to the job. For many jobs, formal written or actual work sample tests may not be a viable alternative, at times simply because there are very few candidates and the cost of test development does not warrant the effort. In these instances, an interview may be the only test in the examination (except the minimum qualifications). For other jobs for which there are many tests in the examination, those responsible for examining have the added obligation of creating and implementing an interview procedure that is well integrated with other tests in the process. Interview materials are typically developed directly from information contained in the job analysis. A viable set of questions for the interview and scoring criteria must be established. In addition, the most effective interview programs include careful standardization of interview administration, a written guide to conducting the interview, and a training session for interviewers. In the public sector, an interview panel is virtually always used as opposed to a single interviewer (or even sequential interviews). Although an interview panel introduces more costs in terms of time and personnel, it has the distinct advantage of enhancing the reliability of the process, and, most importantly, it provides a way of documenting that reliability along with a greater appearance of fairness.

The American Public Transportation Association (APTA) has developed a Bus Operator Selection System (BOSS) that includes a 75-item survey of attitudes, beliefs, and experiences, followed by a multifaceted interview designed to be conducted by a panel of three interviewers, each representing a different perspective on the job: operations, training, and human resources (HR). This system is being used in about 30 transit organizations and has been administered to more than 160,000 candidates nationwide. The original work documenting the system is described in Jacobs, Conte, Day, Silva, and Harris (1996) and highlights the utility of multiple performance predictors for selecting bus operators.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

Alternative Measures

All employers should identify appropriate predictors for use in employee selection and, in addition, are required to search for alternative predictors for any original predictor that demonstrates a marked difference in pass rates on the basis of designated candidate demographic/cultural group membership. For example, cognitive ability tests are effective predictors of subsequent job performance, but the use of cognitive ability tests alone will also usually result in large racial/ethnic group differences in pass rates, with majority group members passing at a higher rate than members of most minority groups. In this case, employers are required to seek out other predictors that can be used in conjunction with or in place of the test(s) that result in large group difference(s) in pass rates. The search for alternatives may take the form of identifying additional dimensions upon which to assess candidates. This approach is reflected in the previously mentioned police officer selection example, in which, for many decades, police candidates were given a cognitive ability test for the initial identification of qualified candidates and then further vetted via interviews, physical ability tests, background investigations, and medical/psychological evaluations. More recently, systems for initial screening commonly include a cognitive test with additional areas measured, such as personality or biographical data. These types of assessment systems can also include other testing formats such as video-based tests or job simulations. Both of these approaches expand test content and format and should be considered when meeting the mandate of alternative tests. Recently, a group of forward-thinking psychologists at Shaker Consulting Group has pioneered an expanded set of predictors they refer to as "The Virtual Job Tryout." These systems have been built for specific clients to capture the complexities of jobs using a variety of types of selection tools (see http://www.shakercg.com). By combining more traditional tests and surveys with video based scenarios requiring preferred response a more complete picture of each candidates' overall job fitness emerges.

Risks and Legal Challenges

Ultimately, any testing system must have a formal evaluation regarding its ability to accurately select individuals. Public sector testing programs are often the first to be challenged because they require use of formalized testing and they impact large numbers of applicants to jobs that are so visible and pervasive in our society. The Equal Employment Opportunity Commission (EEOC), the U.S. Justice Department, and state and local fair employment agencies often scrutinize these highly visible testing programs, and the Uniform Guidelines (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice, 1978) help define the process.

Any test used for selecting a few successful candidates from a large number of applicants is a likely target for challenge, and those responsible for the testing process must have some way of demonstrating its links to the job, its ability to select the right people, and, possibly, why this test was used whereas others (alternatives) were not. This sets a high standard, and one that is required by law. It also demands a strong logical basis for decisions made in developing the testing process and, ultimately, leads to higher-quality tests and more capable individuals being selected. Even with all of these objectives met, those who believe the results were unfair have the right to and can challenge a civil service examination.

CREATING A TALENT PIPELINE: RECRUITING CANDIDATES

No selection program can be successful unless the number of candidates exceeds the number of positions available. This has been an operating principle of employee selection for years and was first codified in the work of Taylor and Russell (1939). The challenge that faces public sector employers is two fold: (a) to create efficient recruitment and selection systems for all jobs to maximize the number of qualified applicants and (b) to expend extra effort to find candidates

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

for jobs where demand has outstripped supply. With respect to the first challenge, although a larger number of candidates is generally seen as a positive in terms of selection utility, assessing very large numbers of candidates can result in much higher selection costs, thereby driving down the overall utility of the program. A case in point is a large police force in New York State. A test is administered once every four years, and approximately 50 new officers are hired each year (about 200 hired over the life of the list). Because this is a particularly attractive job in terms of prestige, location, and salary, the number of candidates is very high. In 2015, there were over 17,000 test takers (in previous years it has exceeded 30,000). Clearly, a selection ratio of 1 in 85 is well beyond what is needed for effective selection and simply increases the cost of testing. In this case, the testing process required use of more than 150 schools and 3,000 test monitors and administrators, and its total cost was estimated to exceeded $2 million (EB Jacobs, LLC, personal communication, June 2015).

In contrast, finding sufficient applicants is the major issue for several public sector jobs. For some jobs, there may be as many openings as candidates and, in some cases, fewer qualified individuals than positions available. When this occurs there is no selection taking place, and the focus must turn to recruiting. The job analysis identifies the KSAOs of people to attempt to attract, making it possible to target the recruiting effort. Prior hiring information can also help identify schools, vocational programs, and other training grounds where those who have been successful in the job were located. Establishing partnerships with educational institutions may also be considered; at times, completion of such a program has even become an MQ (minimum qualification) for the examination. Not all recruiting efforts return positive results. Recruiting without information on the KSAOs required to be successful is likely to be a waste of time and effort. In some cases, "random recruiting" may actually detract from the goal of more efficient testing. When recruits who are poorly prepared or not truly interested in the job are solicited and become candidates, the potential for changes in the passing rates of the various demographic applicant groups increases. The result can be a higher rather than a lower level of adverse impact, no doubt the opposite of the desired outcome.

Indeed, effective recruitment is vital to creating an effective selection program, but the process must be guided by the knowledge of what is required by the job, what have historically been successful avenues for finding the needed talent, and what new approaches (e.g., pre-training programs and educational partnerships) may prove viable in the future.

PROMOTIONAL PROCESSES: USING WHAT WE KNOW ABOUT PEOPLE AND THEIR CAPABILITIES TO OUR ADVANTAGE

A word about some differences between public sector entry-level testing and public sector promotional testing is important to fully understand the various approaches to selection that are required. For many entry-level positions, a formal training program exists, and those selected for the position will be placed in that program once they formally accept the job. In many public sector promotional systems, the individual who was at a lower-level job on Monday may find herself at a higher-level job on Tuesday with little or no training prior to moving up the organizational ladder. This distinction has important implications for testing. Because training will occur for the first example and not for the second, it means that testing for the lower-level position should not include the knowledge, skills, and expertise that will be learned prior to moving into the position. In the promotion situation, the requisite information, skills, and expertise are needed on day one of the higher-level position incumbency, so it is legitimate to test for all. In practice, what this often means is that entry-level tests focus on underlying abilities requisite for learning the job, whereas promotional tests are more closely linked to actual job requirements.

There are also distinctions in terms of the validation processes most often encountered when it comes to entry-level versus promotional testing. A content strategy for validation is likely to be used for either entry or promotional testing. As stated earlier, this method of validation establishes logical links between test requirements and job requirements, often supported with judgment data from subject matter experts (SMEs). In many programs of validation for entry-level testing, this strategy is supplemented with a criterion-related validity study or by citing

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

evidence of generalized validity. However, it is rare that a criterion-related study is part of the validation process in promotional exams. This happens for various reasons, including relatively small sample sizes for many promotable positions, difficulties in motivating current incumbents to sit for an "experimental testing session," and issues of test security that arise as a function of administering a test to a group of incumbents and then using it again for candidates.

The number and variety of *promotional examinations* are also daunting, and the task of creating and implementing job-related examinations is equally difficult. In the vast majority of instances, promotional examinations include minimum qualifications that specify the lower-level class or classes from which promotion to a given class must be made, as well as the required number of years of service in the lower-level class(es). The process of developing a promotional examination is similar to that for entry-level examinations, with one very important difference: Most agencies emphasize promotion from within (again, often legally mandated). As such, the amount of information known about the candidates greatly exceeds what is known about entry-level applicants. This could be a tremendous advantage in the identification of talent if properly tapped.

Developing Promotional Tests

Promotion examinations are developed based on the concept that those in lower-level jobs acquire KSAOs required for the next job in the hierarchy. Similar to entry examinations, potential applicants for promotional examinations are prepared for testing by informing them about (a) the duties and responsibilities of the job, (b) the required knowledge base, and (c) the underlying skills, abilities, and other characteristics required by the job. This information is conveyed to candidates via a test announcement, or "bulletin," which outlines the types of tests, and often their content and scoring, as well as hurdles (decision points for progression in the examination) that candidates will encounter. These promotional processes can range from a single knowledge-based test to very elaborate, multistage assessments involving simulations and assessment center exercises that unfold over a long period of time. For some positions, this may require very little preparation, but for others (e.g., police sergeant or fire captain), agencies often announce the examination six months or more in advance to give candidates adequate time to prepare for the various tests that make up the promotion process.

Appraising Past Performance

One frequently missing element in promotional processes is the assessment of past performance. Although this has the potential to be the most important single indicator of future performance, its rare use in promotional processes stems from a lack of confidence that performance ratings have been or will be consistent and accurate. Concerns of bias in the ratings by supervisors abound. More generally, it is typically believed that performance ratings lack the psychometric rigor required for any formal testing process.

Indeed, one clear opportunity for improving promotional processes is the more effective use of past performance for determining who will move up in the organization. To this end, several assessment techniques, some of which have been used in private sector selection and, especially, in employee development programs, have been devised for the measurement of past performance. Behavioral accomplishment records (Hough, 1984), ratings of "promotability," career review boards, and behavior-based interviews have all been seen as additions to the overall promotional processes used in public sector testing. It remains the task of testing professionals to further enhance promotional processes by continuing to improve these techniques that capture prior job-relevant performance. An often expressed sentiment of promotion candidates is: "The examination should not evaluate what you do on that one test day. It should assess what you do the other 364 days." Tools like accomplishment records and career reviews attempt to incorporate this perspective.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

PERSONNEL DECISION MAKING AND LEGAL JEOPARDY

As noted above, the promotion process (as well as the selection of new employees) in public sector organizations can often lead to legal disputes and challenges by individuals, groups, and government entities, such as the U.S. Department of Justice. Most of the time what is at issue is disparate impact, in which the results of the selection or promotion systems appear to disadvantage one or more demographic/cultural groups. When this occurs, as for private employers, the public sector agency is required to demonstrate the validity of the process. This demonstration can take many forms, and it is not unusual to provide multiple sources of validity evidence, ranging from the most common form, content-based evidence of validity, to extensive documentation of criterion-related validity, which may be based on research conducted internally or by external consultants for the organization and/or generalized evidence of test validity, although "the jury is still out" regarding the degree to which validity generalization has been seen as acceptable by courts and regulatory bodies.

Unique Competitive Processes

The stakes in public sector promotional testing can be very high. As stated above, in many public sector jobs, the only way to advance is by having served in one or more specific job(s) for a minimum number of years, sometimes additionally having successfully completed specialized education/training or other formal certification, and by successfully competing in the promotional examination process. In these examinations, some candidates succeed, but a larger number of candidates do not. This direct competition among peers can have negative consequences for the individuals involved and for the organization. The entire process may be challenged by individuals who did not do well enough to be promoted and have thereby concluded that the process was flawed and unfair. When this happens, colleagues find themselves on opposite sides of a legal battle, in which the candidates who were successful during the testing process hope the test results will be upheld, and those who did not do sufficiently well on the examination to be promoted work to discredit the process. Unfortunately, most candidates have often invested a great deal of preparation time, and many feel frustrated by the delay in implementing the results. These types of challenges may stretch out for years, creating problems for all participating entities: the candidates, the HR professionals, and management of the agency wishing to promote its employees.

Another factor that affects the tendency for legal challenges of selection processes within the public sector, in contrast to much of the private sector, is that these processes are by design open and visible; unquestionably, the examination is "responsible for" selection outcomes. This provides disappointed candidates an obvious target for pursuit of litigation. Furthermore, because civil service systems still very frequently have mandated candidate appeal or "protest" rights, filing a lawsuit may seem nothing more than an obvious extension of a right they are already afforded. In point of fact, formal, stringent requisites of what constitutes an appeal or protest and how they are adjudicated exist, but these rights at times seem to be misinterpreted simply as a right to register complaints. Once administrative remedies have been exhausted and the outcome remains negative to the candidate's interest, it may seem a natural next step to pursue litigation.

Negative Consequences for Individuals and Organizations

The legal challenges that at times confront public sector testing may create a crisis of confidence in testing. Individuals may begin to question the ability of the people responsible for testing and speculate that the system has come under the control of the legal system without regard for merit. As these cases drag on, temporary appointments may be made, which further complicate the situation. When order is finally restored, another problem can occur with respect to what to

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

do with those who were placed in the higher-level job as a provisional appointment. When a new testing program is instituted and someone who has been in the job for many months or even years does not achieve a successful score, the immediate question that arises is "How could the test be relevant to the job (valid) if someone who has managed to do the job successfully for the past few months/years cannot pass it?" In this context, no consideration is given to actual job performance, and those who have been successful for months or years can be taken out of that job based on the results of a day or less of testing. There exists no simple or single answer to this dilemma.

Balancing Validity and Diversity

Public sector agencies are in a constant struggle to simultaneously increase the validity of their selection and promotion processes and to improve the diversity of the group that is selected. This complex task may involve actions that result in focus on one of these objectives at the expense of the other (Aguinis & Smith, 2007; DeCorte, Lievens, & Sackett, 2007; Lindsey, King, McCausland, Jones, & Dunleavy, 2013; Ployhart & Holtz, 2008; Pyburn, Ployhart, & Kravitz, 2008; Sackett & Lievens, 2008). Many of the tests used to predict future performance show large differences among various groups. With respect to a variety of measures of cognitive ability and knowledge-based multiple-choice tests, both of which are popular with public sector agencies because of their clear right-and-wrong response format, Caucasian candidates consistently outperform Black and Hispanic candidates. When the selection procedure switches from cognitive ability to physical ability, women typically score lower than men. Agencies take steps to minimize these differences, and although some approaches may be helpful, (e.g., practice testing), none eliminate the group differences completely, and some practitioners argue that those who already have what it takes just get stronger.

Recently, many public sector agencies, although still acknowledging the need for some component of the process to test for cognitive ability, have created examinations that include noncognitive measures such as personality tests or biographical information. These types of tests are sometimes met with protest from applicants, unions, and other interested parties on several grounds, but, at least when it comes to selection of new candidates (versus promotional testing), such testing has been implemented. In some instances, the inclusion of different types of instruments has reduced group differences that are observed when testing only for cognitive ability and has also enhanced overall validity, but they have not eliminated adverse impact (Sackett & Lievens, 2008). In some instances there is a reduction of group differences by changing the weighting of the various test components (DeCorte, et al., 2007; Decorte, Lievens and Sackett, 2011). One of the reasons for this failure in eliminating adverse impact is a low but consistent correlation among cognitive-ability-oriented predictors and less traditional selection tools such as personality indicators. Although many personality scales show no difference between minority and majority group members, our work with police officers and firefighters has shown that some scales do show differences in the context of public safety selection. The difference is also in favor of majority test takers in a way that is believed to be linked to the positive correlation between these personality measures and cognitive ability and in a manner that inhibits their ability to reduce adverse impact (Cascio, Jacobs, & Silva, 2010).This problem is made even more difficult by the fact that, for many public sector jobs, the selection ratio is quite favorable for the organization (i.e., many candidates and few individuals selected). As the selection rate gets smaller and smaller (more candidates relative to the number of positions to be filled), the impact of any group difference grows quickly. Even small group differences can cause large levels of adverse impact when selection ratios drop below .20. This further complicates the situation for the agency, because one goal is to make the jobs widely available, but doing so can have a negative consequence on diversity. Important to recognize here is the fact that as selection rates go down, it is often because the number of applicants is quite high. With large numbers of applicants, statistical tests for adverse impact move in the direction of an increased finding of adverse impact. This relationship between the size of the applicant pool, selection ratio, and adverse impact has been highlighted by Jacobs, Murphy, and colleagues (Jacobs, Deckert, & Silva, 2011; Jacobs, Murphy, & Silva, 2012; Murphy & Jacobs, 2012).

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

Defensibility of Process

Ultimately, a public sector agency must make its employee selection systems (entry and promotional) defensible. To do this, steps must be taken, and these steps must not only conform to the laws and guidelines governing selection, but they must also be meticulously documented (Guion, 1998).

In entry and promotional selections, there are winners and losers. The winners get what they desired, the job, and those who are less fortunate walk away either without a job (entry) or in the same job they were in prior to the promotional process. At times, those in the latter category seem to decide that challenging the test is a means of obtaining a second chance. In some situations, the tests may actually be poorly prepared, lacking in job relevance, undocumented with respect to how they were created and/or linked back to the job, or simply administered without regard to accepted testing practices. However, in other cases, the allegations about the test may be a disingenuous attempt to vacate the results and provide all test takers with another opportunity for success. When a test is challenged, it does not automatically mean the test was deficient or that the process violated the laws and guidelines that prevail. Challenging the test is a right of any candidate who can establish an underlying legal basis, most often in the form of adverse impact. Once adverse impact is established, it becomes the responsibility of those using the test to establish the validity of the process.

Given the above, it is important to consider what must be present to make a hiring or promotional system defensible. Below we provide further details on the following critical factors for defending a testing process:

- Job analysis
- Links between test elements and aspects of the job
- Logic behind the combination of multiple test scores into a final composite
- Test administration details
- Scoring processes
- Documentation of the entire testing program from start to finish

There is unanimous agreement that a fair test is only possible with confirmation that those responsible for the test understand the job. In the context of public sector testing, this means that the job in question has been defined and documented, which occurs at the inception of a job class through the creation of a class specification, and then is often supplemented with a more detailed job analysis and/or during examination development with the assistance of job experts. The results are widely accepted by incumbents, supervisors, HR specialists, and potential applicants as reflecting the important features of the job. To be useful as underlying test development documents, the class specification and job analysis must reflect not only the tasks and responsibilities of the job but also the knowledge base required by the job and those skills, abilities, and other personal characteristics that facilitate job performance.

A second important element of providing the necessary information for defense of a test is evidence that links the test items, work samples, and other components of the examination to the tasks performed on the job. It is helpful to think of this in one of two ways. The most direct way can be seen in work-sample testing, where the test is actually a sample of the tasks performed on the job. Physical ability testing provides the best example of this direct linking. Firefighter candidates are often asked to perform a series of job-related activities such as advancing a fire hose, dragging a dummy, and climbing stairs with equipment. In the case of each event included in the test, the activity is a replicate of what is done on the job. One challenge here is to make sure that key skills learned during training are not required in the performance of test events. Recently, a test was developed for store clerks for a retail drug store chain. Clerks repeatedly engage in unloading boxes and in stocking shelves. These job activities require little in the way of specialized training, and a test was developed to replicate the unloading and stocking tasks. Here there is a direct link between the job and the "items" on the test. More often this logic requires two linkages: the first relates the tasks and requirements of the job to a set of knowledges, skills, and abilities, and then a second linkage is required to show the relationship between test items/elements to these same KSAs.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

This *linking process* often takes the form of surveys that identify the knowledge or other attributes underlying the test questions and asks job experts to identify the degree to which each aids in completion of various tasks. This process is commonly accepted as a means of establishing validity on the basis of test content. At the root of any successful demonstration of validity is a clear listing of test items, the job tasks, and the "linkage" of the two. Critical to this approach to the demonstration of validity based on test content is an appropriate sampling of incumbents, supervisors, and/or other job experts, along with clear instructions to those who are providing the responses. Although surveys are often used, this process can also be accomplished with review meetings involving job experts in which the material is analyzed and discussed and consensus judgments of the experts are documented.

In most contemporary entry and promotional processes, a single test does not represent the full range of job requirements, so multiple tests are used for selection or promotion. Yet a single, final score in the examination is required and, therefore, the manner in which that score is calculated becomes an important consideration. Selection is often based on a written knowledge- or ability-based test and a series of interviews. The process requires that a list of candidates is created, and to do this these different assessments must be turned into a composite score. The logic of how the composite is formed can be taken from job analytic information that indicates the degree to which each score is related to job KSAs, the reliability of the score, and the overall importance of that indicator to job performance or other measures that provide a logic regarding aggregation of information. (Composite predictor scores are addressed in more detail in Chapter 17, this volume.)

Another and perhaps more complex example can be seen in police officer selection and in firefighter selection, where there often is a written and a physical test. Combining these two scores becomes an issue because the weight assigned to each score will determine, in part, its impact on final score. Years of job analysis for both of these jobs have yielded consistent results. Although both jobs require physical capability, the firefighter job is more physically demanding. Results from our own job analysis work across various police and fire departments have shown that the job of a firefighter is between 40% and 60% physical, with the remainder requiring cognitive abilities, whereas the job of a police officer is often reported by incumbents to be between 20% and 30% physical and 70–80% cognitive. The two jobs clearly need different weights for the physical test when it comes to creating a selection composite.

Like the evidence for linking the test elements to the job requirements, a rationale for the weighting used to form the final test score composite is necessary. This is often based on input from job experts and professionals in the testing area. The important point is that the rationale is tied to requirements of the job. There is no one best way to establish these weights; also, in many testing situations, the components of the examination are correlated with one another. When correlation exists among components, the weights become somewhat less of an issue because small variations in weights do not substantially change the overall results. As the correlations increase, the impact of differentially weighting the components becomes far less of an issue, and at some point, simply equally weighting the test components works in a similar manner to elaborately defining a very precise set of differential weights. On the other hand, some would argue for use of equal weights simply because of their demonstrated robustness in prediction (Schmidt, 1971). Either way, differential versus equal weights, there is a need to standardize each test score used so that the effective weights approach the intended weights. Without standardization of scores, the tests with the larger variations will have larger contributions to the total score.

A fourth area for defensibility is in the actual administration of the tests. The best-developed tests, the ones with the highest degree of validity evidence and the strongest rationale for weighting of components, can become useless if test administration processes are deficient. Threats to administration can come in various forms, ranging from failure to protect testing materials before the actual test date to administering the test in a room with poor lighting, loud outside noises, or missing pages in test booklets. Although this seems to be the least difficult part of defending a test and the easiest to achieve, it is often the Achilles heel of a testing process. Care must be given to all phases of test administration; for example, materials such as instructions to candidates, information about the testing locations and facilities, and any irregularities in the actual administration of the test all must be well documented. Otherwise, one may do all

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

of the right things when it comes to test development, but then compromise it all during test administration.

All test materials must be scored, and the scoring process represents a fifth area in which threats to the defense of a test can occur. Many modern tests are administered via paper and pencil and scored by scanning machines or taken online and scored automatically. Scoring integrity must be demonstrated in the form of getting the correct outcome for each candidate. In the case of scanned answer sheets, this means that all answer sheets must be reviewed for irregularities. It is a good idea to scan each test sheet twice and to compare scores for any differences in the two scans; any differences indicate scanner problems or simple "hiccups" in the scanning process. Another way to ensure accuracy is to compare candidates' hand-scored tests with their scanned scores (usually for a sample of candidates). For online testing, periodically sending through a "phantom candidate" with a known score to make sure that the algorithm is generating the correct score is a useful step. With respect to other types of potential test scoring problems, demonstrations of inter-rater agreement and other forms of reliability help to substantiate the appropriateness of scoring protocols. Although a discussion of reliability is not consistent with the goals of this chapter (see Chapter 1, this volume), any and all steps that can be taken to show the consistency of test results will be of great assistance in addressing any challenges to the scoring process.

A final step in the defensibility of a testing program is the documentation of all steps in the process. This includes specification of how each test was developed, how each test was administered and scored, and how the final examination score was calculated for all candidates. Creating the paper trail of your work not only allows everyone to see the steps taken but also memorializes the process. In many challenges to public sector testing, legal proceedings take place years after the examination was developed and administered. Relying on memory and randomly filed memos of what happened will never provide the information necessary to successfully support the contention of adequacy of the process. Public sector testing is best completed by the compilation of a final report or file that details the project from start to finish. This documentation should be clear, and it should contain all of the necessary surveys, instructions, and tests used in the examination. There is no better way to defend a test than to have it well documented. In situations in which a challenge is presented to an agency regarding the testing process, the agency can provide the potential plaintiff with a copy of the report. On more than one occasion, this has ended the challenge to the test.

CONCLUSIONS

Public sector testing has evolved over the past two centuries in terms of test content and test format. We have seen the movement from tests based solely on memory and other cognitive abilities to the inclusion of social judgment, personality, and biographical information. We have seen simple paper-and-pencil testing transition to testing formats that are computer-based and inclusive of video stimulus materials.

It is our sincere wish that those involved in public sector testing continue to engage in the scientific-practitioner model, where innovations in assessment that appear in the research literature are incorporated into testing programs and that progressive testing programs are described in our research journals. One example of such actions is the move to provide assessments of cultural proficiency during the hiring process for police officers. Much has been written about racial bias in policing, and the news has documented many cases where police response has been lethal and beyond what is considered reasonable. Much has been written about the assessment of racial bias, and now police departments are pushing to include a mechanism for measuring racial bias during the selection process. Success in this area should be further documented so others can benefit from the application of scientific knowledge to real-world problems.

With respect to the identification of critical underlying job requirements, we have seen public sector testing programs expand their use of systematic job analytic techniques that approach not only single jobs under study but also groupings of jobs, so that the inherent interrelationships among jobs can be identified to better take advantage of opportunities to use a common testing

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

system across jobs. With respect to the legal arena, public sector testing is often singled out as the test case for looking at the defensibility of specific test formats and test content as well as the way in which test scores are used in the decisions made about people and jobs. Clearly, as the challenges to the fairness of various types of testing programs move forward, public sector applications will be part of the landscape.

Unlike the private sector, public sector employees are less susceptible, although still not immune, to layoffs or downsizing, although hiring freezes are common. This fiscal reality translates to the fact that most cities and public agencies, even in their toughest financial times, continue to require substantial levels of staffing. Therefore, the enormous demand for testing programs for the hiring and promotion of public sector employees will continue, and the need for accomplished and creative test development professionals will offer tremendous opportunities to further develop the way in which we measure candidates against job requirements.

REFERENCES

Aguinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology, 60*, 165–199.

Baruch, I. (Chair, Committee on position-classification and pay plans in the public service.) (1941). *Position-classification in the public service*. Chicago, IL: Public Personnel Association.

Cascio, W. F., Jacobs, R. R., & Silva, J. (2010). Validity, utility and adverse impact: Practical implications from 30 years of data. In J. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 271–288). New York, NY: Psychology Press.

City of Los Angeles. (2009). *Official city of Los Angeles charter. American Legal Publishing Corporation*. Retrieved November 16, 2009, from http://www.amlegal.com/nxt/gateway.dll?f=templates&fn=default.htm&vid=amlegal:laac_ca

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology, 92*, 1380–1393.

De Corte, W., Lievens, F., & Sackett, P. R. (2011). Designing Pareto-optimal selection systems: Formalizing the decisions required for selection system development. *Journal of Applied Psychology, 95*, 907–926.

DuBois, P. H. (1970). *A history of psychological testing*. Boston, MA: Allyn & Bacon.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*, 382990–38315.

Franke, W. (1963). *The reform and abolition of the traditional Chinese examination system*. Cambridge, MA: Harvard University Press.

Graham, J. R., & Lily, R. S. (1984). *Psychological testing*. Englewood Cliffs, NJ: Prentice Hall.

Gregory, R. J. (1996). *Psychological testing: History, principles and applications* (2nd ed.). Boston, MA: Allyn & Bacon.

Hollenbeck, G. P., McCall, M. W., & Silzer, R. F. (2006). Leadership competency models. *The Leadership Quarterly, 17*, 398–413.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum.

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

Jacobs, R. R., Conte, J. M., Day, D. V., Silva, J. M., & Harris, R. (1996). Selecting bus driver multiple perspectives on validity and multiple estimates of validity. *Human Performance, 9*, 199–218.

Jacobs, R. R., Deckert, P. J., & Silva, J. (2011). Adverse impact is far more complicated than the uniform guidelines indicate. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*, 558–561.

Jacobs, R. R., Murphy, K., & Silva, J. (2012). Unintended consequences of EEO enforcement policies: Being big is worse than being bad. *Journal of Business and Psychology, 28*, 467–471.

Lindsey, A., King, E., McCausland, T., Jones, K., & Dunleavy, E. (2013). What we know and don't: Eradicating employment discrimination 50 years after the Civil Rights Act. *Industrial and Organizational Psychology*, *6*, 391–412.

Murphy, K., & Jacobs, R. R. (2012). Using effect size measures to reform the determination of adverse impact in equal employment litigation. *Psychology, Public Policy and Law Journal, 18*, 477–499.

Murphy, L. L., Plake, B. S., & Spies, R. A. (Eds.) (2006). *Tests in print VII*. Lincoln, NE: Buros Institute of Mental Measurement.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*, 153–172.

Pryburn, K. M., Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61*, 143–151.

Rodrigues, D., Patel, R., Bright, A., Gregory, D., & Gowing, M. (2002). Developing competency models to promote integrated human resource practices. *Human Resource Management (Special Issue: Human Resources Management in the Public Sector), 41*, 309–324.

Sackett, P., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419–450.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement, 31*, 699–714.

Spies, R. A., Plake, B. S., & Geisinger, K. F. (Eds.) (2007). *The seventeenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurement.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology, 23*, 565–578.

Downloaded By: 10.2.97.136 At: 15:40 26 Sep 2023; For: 9781315690193, chapter32, 10.4324/9781315690193-32