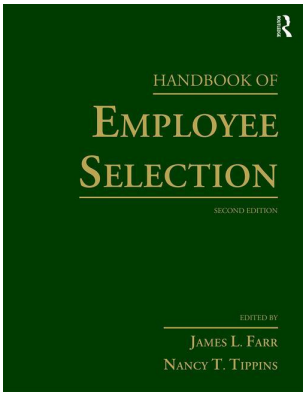

Handbook of Employee Selection

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

An Updated Sampler of Legal Principles in Employment Selection

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-28
Arthur Gutman, James L. Outtz, Eric Dunleavy
Published online on: 22 Mar 2017

28

An Updated Sampler of Legal Principles in Employment Selection

ARTHUR GUTMAN, JAMES L. OUTTZ, AND ERIC DUNLEAVY

INTRODUCTION

We are very pleased that we have been asked to contribute an updated chapter on legal principles for this volume.[1] As we noted in the original chapter, many principles of employee selection have been illuminated through decades of equal employment opportunity legislation, regulation, and related case law. Granted, judges write opinions, but these judges frequently depend heavily on industrial-organizational (I-O) psychologists in framing these opinions and are generous in their citations to our work and our testimony. Once again, we see a clear role for employment litigation knowledge related to personnel selection in this volume.

In the context of this role, we have been asked to review, refresh, and expand upon the "principles" that were established in the original chapter. We focused those principles on selection practices that are commonly at issue in employment discrimination cases. We intended the "principles" to represent general themes that practitioners might keep in mind when attempting to minimize legal risk when using such practices and identified exemplar court cases where we could.[2]

We use the term "principle" to refer to a basic and generally accepted combination of science and practice as they relate to specific legal matters. This use is obviously different from "the" *Principles* promulgated by the Society for Industrial and Organizational Psychology (SIOP; 2003), which were being revised for a fifth edition at the time this book chapter was written. The SIOP *Principles* are intended to be a scientific document and not a legal one, whereas the principles we denote in this chapter are focused on the equal employment opportunity (EEO) legal context. Neither are we intending to address the "best" practice from a theoretical or empirical perspective. Instead, we are simply articulating what might be seen as contemporary "defensible" practice from the legal perspective. Nevertheless, the principles we present are general rather than specific and are generally well documented via a variety of sources.

Like other chapter authors in this Handbook, we are constrained by allotted pages. Therefore, we have been selective in formulating principles and providing underlying case law. As a result, we do not consider these principles to be exhaustive, or possibly even the most important, from a psychometric or theoretical sense. Although we would consider these principles more rather than less important than others that might have been chosen, we acknowledge that there may be other equally important principles that we might have chosen to articulate: so many principles, so few pages.

The principles we have chosen to present generally deal with either a general practice (e.g., work analysis, adverse impact), a statute (e.g., ADA, CRA, ADEA), a protected group (e.g., gender, race, age, disability), or some combination of those factors. The employment selection

practices we have chosen to address represent "terms and conditions" of employment as defined in key federal laws. Terms and conditions are broadly defined to include recruitment, hiring, training, promotion, benefits, and termination. The term "law" is broadly defined as including statutes, executive orders, and constitutional amendments. Broader coverage of EEO law is provided by Gutman, Koppes, and Vodanovich (2010), Landy (2005), and Outtz (2010). As a group, these EEO laws proscribe discrimination in the workplace on the basis of race or color, religion, sex, national origin, age, and/or disability.

In some ways, Title VII of the Civil Rights Act of 1964 is the most critical of the statutes. It includes proscriptions relating to the largest number of protected classes (race, color, religion, sex, and national origin) and the full gamut of terms and conditions of work. Title VII also provides a critical model for later statutes, including the Age Discrimination in Employment Act of 1967 (ADEA) and the Americans with Disabilities Act of 1990 (ADA) as amended by the ADA Amendments Act (2008), which address the full gamut of terms and conditions for age and disability, respectively. The ADA updates the Vocational Rehabilitation Act of 1973 (Rehab-73). Section 503 of this Act was updated in March of 2014 (as well as the Vietnam Era Veterans' Readjustment Assistance Act of 1974), and this change is covered in Principle 14. Finally, the Equal Pay Act of 1963 (EPA) is a more surgical statute that proscribes only wage discrimination on the basis of sex, which is also included in Title VII as part of terms and conditions.

Among other relevant laws, Executive Order 11246 (EO 11246) contains rules for contractors doing business with the federal government. These contractors must agree to provide (a) equal employment opportunity for all groups and (b) affirmative action for minorities and women if there are discrepancies between their representation in the employer's workforce and qualified workers in the labor pool. It is critical to note that the primary requirement for affirmative action in EO 11246 is recruitment and outreach, and there is no requirement (in fact, it is illegal under Title VII and other laws) to make actual selection decisions on the basis of race or gender.

The key constitutional amendments governing employment decisions are the Fifth and Fourteenth Amendments, which apply to federal and state employers, respectively, and which have been used extensively in so-called reverse discrimination lawsuits on the basis of race or sex, providing overlapping coverage with Title VII. Additional overlapping coverage with Title VII is represented in the Thirteenth Amendment, which applies to the full range of terms and conditions of employment as in Title VII, with the exception of adverse impact challenges. Critically, although each of these amendments preceded Title VII, no one ever thought to use them in workplace discrimination cases until after Title VII, not only for overlapping coverage but also for gaps in coverage in Title VII that do not apply to constitutional claims, such as the minimum number of employees required in Title VII ($N = 15$); no such minimum workforce size exists for constitutional claims.

Some final but important points to note are that each of the statutes, except for Rehab-73, requires that claims of discrimination be processed by the Equal Employment Opportunity Commission (EEOC). In comparison, constitutional claims can go directly to federal district court. Additionally, the Office of Federal Contract Compliance Programs (OFCCP) of the Department of Labor (DoL) regulates EO 11246 and Rehab-73 in the nonfederal sector and typically uses an audit system for investigation. EEOC regulates EO 11246 in the federal sector, and the EEOC and Merit Systems Protection Board (MSPB) share responsibilities for federal claims under Title VII and the ADEA.

Because this entire volume is dedicated to the concept of "selection," some of our principles may appear to stray from that mark. We construe selection broadly to encompass many personnel actions. These include, but are not limited to, applicant screening, hiring, promotion, selection to training experiences that are gateways to promotion, and layoffs (or deselection). We believe that selection should be broadly and not narrowly construed because in each case a personnel decision is being made that implies a direct or indirect movement (or lack thereof) of an applicant or a current employee. Finally, in the context of a selection or promotion decision, we will occasionally comment on a co-relevant issue such as compensation or training. It is not that we intend to turn the focus to compensation, which has been an interesting EEO focus[3] under the Obama administration, but rather that compensation decisions are often made at the same time as selection or promotion decisions and deserve comment.

PRINCIPLES AND EXEMPLAR CASE LAW

1. Companies using Internet recruitment should understand the definition of "Applicant" and be prepared to defend qualifications listed in the job description as well as procedures used to screen applicants.

After publication of the *Uniform Guidelines on Employment Selection Procedures* (UGESP) in 1978, the EEOC answered a series of follow-up questions in 1979 (44 FR 11998). Among them, Q15 defined the term "Applicant" as follows:

> The precise definition of the term "applicant" depends upon the user's recruitment and selection procedures. The concept of an applicant is that of a person who has indicated an interest in being considered for hiring, promotion, or other employment opportunities.
>
> (UGESP, Q15)

This definition was rendered obsolete by subsequent developments in Internet recruitment, because it became easy to indicate interest in many jobs at the same time. Therefore, the EEOC amended Q15 on March 4, 2004 (FR Doc 04–4090) to require more than an expression of interest. Additionally, the OFCCP issued separate guidance on March 29, 2004 (41 CFR Part 60–1).

The OFCCP focuses on broad recordkeeping for EO 11246, whereas the UGESP focuses on adverse impact under Title VII. Readers should note that SIOP's Professional Practice Committee provided a detailed evaluation of the EEOC and OFCCP documents (Reynolds, 2004).

The EEOC answered five new questions. Two of the answers clarify that Internet methodologies are covered by laws such as Title VII and by the UGESP, and two others clarify that Internet search criteria are subject to adverse impact rules and that selection procedures administered online are subject to the UGESP. The last and most critical of the five answers defines the term "Applicant" using the following three prongs:

1. The employer has acted to fill a particular position.
2. The individual has followed the employer's standard procedures for submitting applications.
3. The individual has indicated an interest in the particular position.

For prong 1, the EEOC cites a company seeking two hires from among 200 recruits in a database. If 100 recruits in the database respond to employer inquiries about the position and 25 are interviewed, then all 100 responders are applicants and all 100 nonresponders are not applicants.

For prong 2, the EEOC cites two examples: (a) if employers require completion of an "online profile," only those completing the profile are applicants; and (b) if employers e-mail job seekers requesting applications, only those who respond are applicants.

The prong 3 answer clarifies that individuals are not applicants if they (a) only post a resume, (b) express interest in several potential jobs, or (c) follow prong 2 for a job other than those the employer acts on (prong 1).

OFCCP acknowledged the differences in recruiting practices from 1979 to present day in explaining the need for an updated applicant rule. In 2005, OFCCP issued regulations creating the Internet Applicant Recordkeeping Rule ("Internet Applicant Rule"), at the same time acknowledging that the UGESP's definition was overly broad.[4] The OFCCP established four criteria to define an "Internet Applicant":

1. The individual submits an expression of interest in employment through the Internet or related electronic data technologies;
2. The contractor considers the individual for employment in a particular position;
3. The individual's expression of interest indicates the individual possesses the basic qualifications for the position; and
4. The individual at no point in the contractor's selection process prior to receiving an offer of employment from the contractor, removes himself or herself from further consideration or otherwise indicates that he or she is no longer interested in the position.[5]

The addition of basic qualifications is a key distinction between EEOC and OFCCP definitions. Under the Internet applicant rule, basic qualifications must be (1) non-comparative, (2) objective, and (3) relevant to performance of the particular position. Another key distinction is that OFCCP allows for application of neutral "data management techniques" as part determination of whether someone was considered for employment. Clearly, differing definitions of applicant could meaningfully affect who is included and excluded from analyses, which in turn could influence the results of adverse impact analyses.

Exemplar Case Law Related to Principle 1

Although there is as yet no case law citing the new definition of "Applicant," or pitting EEOC and OFCCP definitions against each other, there is a relevant case decided under prior rules (*Parker v. University of Pennsylvania*, 2004). Parker, a White male, filed a Title VII reverse-discrimination claim for failure to consider his application for various jobs at the university. Parker submitted his resume online. The university's website advised individuals to submit their resumes into the database and, if they so choose, to apply for specific job postings. Parker received a letter stating that, if appropriate, he would receive a letter within 30 days notifying him of an interview, but that "otherwise this will be your only communication from us." Summary judgment was granted to the university because Parker never expressed interest in a specific posted job. Parker also claimed adverse impact, but it was dismissed because he could provide no basis for assuming he was harmed. It should be noted that as a standard university policy, recruiters routinely searched the database and forwarded resumes to hiring officers of individuals who expressed interest in a posted job and who met the minimum qualifications (MQs) for that job.

Interestingly, the district court judge ruled that Parker satisfied the four-prong *prima facie* test for disparate treatment from *McDonnell Douglas v. Green* (1973) in that he (a) was a protected class member, (b) applied for a position for which he was qualified, (c) was not hired, and (d) positions remained open for which Parker was qualified. However, the judge also ruled that the university articulated a legitimate nondiscriminatory reason for not hiring him and that Parker could not prove that the articulation was a pretext for discrimination. Accordingly:

> Defendant claims to have failed to consider Parker's application because of its resume reviewing procedures: The reason Penn did not hire Parker is because he had not applied for any specific position, and it did not conduct any search of the resumes of non-applicants for positions being filled.

There are several things to note about this ruling. Although under prior rules, Parker would have been considered an applicant because he had submitted a general application, on the basis of the new rules Parker should not be considered an "Applicant" and therefore should lose at the *prima facie* level. Nevertheless, even nonapplicants may file valid adverse impact claims if they are "chilled" by MQs such as education requirements that disproportionately exclude minorities (*Albemarle Paper Co. v. Moody*, 1975; *Griggs v. Duke Power*, 1971) or physical requirements such as height and weight that disproportionately exclude females (*Dothard v. Rawlinson*, 1977).

2. Adverse impact may be measured using a variety of approaches. Both statistical significance tests and measures of practical significance may be useful approaches to measuring adverse impact.

The measurement of adverse impact is a topic that continues to cause controversy.[6] Some strategies for measuring adverse impact were codified in federal regulations (U.S. Equal Employment Opportunity Commission, 1978), whereas others were imported from foreign case law or can be found in regulatory compliance manuals (Cohen & Dunleavy, 2010). All of these methods have been criticized in the scholarly literature (e.g., Roth, Bobko, & Switzer, 2006).

Statistical significance tests are one approach to adverse impact measurement, and they typically evaluate the hypothesis that the applicant population contains no differences in hiring rates among subgroups. Statistical significance testing has become a preferred method of evaluating adverse impact both by the courts (Esson & Hauenstein, 2006) and by federal agencies such as the OFCCP (Cohen & Dunleavy, 2009, 2010).

These tests can be either directional or nondirectional depending on context and hypotheses of interest. The decision between these approaches can be controversial, because the choice of a one-tailed test can imply that we are only interested in identifying disparities against only one group. Adverse impact statistics are often evaluated using two-tailed significance tests (e.g., OFCCP, 1993), particularly when analyses are proactive. When a specific claim of discrimination is made based on persuasive historical or anecdotal context, a one-tailed test may be reasonable.

A variety of statistical tests that vary slightly based on data system assumptions are available to assess adverse impact. These include the *Z*-test for the difference between two proportions, the Chi-square test of association, Fisher's Exact Test, and Lancaster's Mid-P (LMP) test.[7]

The statistical significance paradigm's appeal is intuitive in that it assesses whether results could be due to chance and provides a yes-or-no answer. However, it is not the only approach available, and in fact the question it answers is fairly narrow. The I-O psychology literature has recently noted the limitations of stand-alone significance tests (e.g., Dunleavy & Gutman, 2011; Jacobs, Murphy, & Silva, 2013; McDaniel, Kepes, & Banks, 2011; Murphy & Jacobs, 2012) and in practice-focused technical guidance (e.g., Cohen, Aamodt, & Dunleavy, 2010). This notion was recently echoed in a statement from the American Statistical Association (2016), which noted that "Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold" and "A p-value, or statistical significance, does not measure the size of an effect or the importance of a result."

In the social scientific realm, assessing the magnitude of variable relations is well accepted. Meaningful relationships between variables are usually more important than trivial nonzero relationships, regardless of significance test results. Yet the challenge is distinguishing what is meaningful from what is trivial. Assessment of practical significance involves both an effect size measure and a standard for determining what is of sufficient magnitude to create concern. A recent technical advisory committee (TAC) on adverse impact analyses (Cohen et al., 2010) noted that practical significance is an important consideration in addition to statistical significance.[8] The TAC noted that it makes particular sense to evaluate practical significance after a disparity has been identified as statistically significant, or unlikely due to chance.

There are many measures of practical significance, including difference measures like the absolute difference in rates and the h statistic difference transform, ratio measures like the adverse impact ratio (often evaluated via the four-fifths rule) and odds ratio, and association measures like the phi coefficient, and phi squared, which is a proxy for traditional variance accounted for measures. The courts have also considered measures of class size as a practical significance consideration (e.g., shortfall, shortfall relative to the number of employment decisions made), as well as flip-flop rules that assess whether a small set of changes in the data affect conclusions (see Morris & Dunleavy, 2015, for more details).

Exemplar Case Law Related to Principle 2

Statistical significance tests first appeared in *Castaneda v. Partida* (1977), which was a jury selection case. Later that year, statistical significance testing was applied in the Supreme Court ruling in *Hazelwood School Dist. v. United States* (1977), where a "*two or three standard deviations*" standard was established.[9] A p-value of .05 (roughly corresponding to two standard errors) means that there is a 5% probability that the observed difference in selection rates for two groups could have occurred due to chance, given a neutral selection procedure. As noted above, significance tests are found in compliance manuals and are generally given the most deference in case law.

The four-fifths rule, which is a standard for evaluating the impact ratio, is described in the UGESP as follows:

> A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.

The four-fifths rule has been endorsed by case law. As an example, the 2nd Circuit ruled in *Waisome v. Port Authority* (1991) that a four-fifths rule violation was required even in situations where a disparity was statistically significant:

> We believe Judge Duffy correctly held there was not a sufficiently substantial disparity in the rates at which black and white candidates passed the written examination. Plainly, evidence that the pass rate of black candidates was more than four-fifths that of white candidates is highly persuasive proof that there was not a significant disparity.

Numerous courts have evaluated practical significance using differences in selection rates. For example, in *Frazier v. Garrison I.S.D.* (1993), a 4.5% difference in selection rates was deemed trivial when most applicants were accepted. In *Moore v. Southwestern Bell Telephone Co.* (1979), where the court held that "employment examinations having a 7.1 percentage point differential between black and white test takers do not, as a matter of law, make a prima facie case of disparate impact."[10]

Several adaptations, or "flip-flop rules," are seen in case law. These rules evaluate the analytic consequences of making slight alterations to the underlying data used for analyses. For example, in *Contreras v. City of Los Angeles* (1981), the number of additional protected-group applicants who would need to be selected to eliminate the statistical significance of the disparity was evaluated. Similar approaches were used in *U.S. v. Commonwealth of Virginia* (1978) and in the *Waisome* case described above. In these cases, if "one or two" additional passes from the "victim" group changed the statistical results, then the disparity was deemed trivial.

More recently, courts have disagreed on how to measure adverse impact. In *Stagi v. National RR Passenger Corp.* (2012) and *Smith v. City of Boston* (2015), judges generally deferred to statistical significance tests over practical significance measures in concluding meaningful adverse impact. In *Apsley v. Boeing* (2013), an appeals court focused more on adjusted shortfall metrics than on statistical significance results and concluded that a disparity was trivial. In *Lopez v. Lawrence* (2014), a judge seemed to consider both significance tests and the four-fifths rule in concluding meaningful disparity.

3. The search for alternatives to procedures that might result in or have resulted in lesser adverse impact is typically split between defendants and plaintiffs. At the outset of a selection project, employers should typically consider alternatives that may meet business needs, be job-related, and minimize adverse impact. After a demonstration of job-relatedness, the plaintiffs have the burden of demonstrating that there is an alternative procedure that would result in the same levels or similar levels of job-relatedness but shows lesser adverse impact.

In years past, for both of these stages (initial research for the employer and post-job-relatedness demonstration for the plaintiff), the search for alternatives may have been considered an onerous and overly burdensome task. However, the I-O literature has expanded significantly since the mid- to late 1990s with regard to decision-making tools and strategies for assessing the tradeoffs between validity and adverse impact.[11] As an example, De Corte, Lievens, and Sackett (2007) investigated the possibility of combining predictors to achieve optimal tradeoffs between

selection quality and adverse impact. This article was a follow-up to earlier research in this area. De Corte (1999) and Sackett and Ellingson (1997) investigated the effects of forming predictor composites on adverse impact. Other researchers have used meta-analytic techniques to forecast the likely outcomes of combining high and low adverse impact predictors in a selection procedure (Bobko, Roth, & Potosky, 1999; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Thus, employers and their consultants (in-house or external) may want to become familiar with this literature to meet their burden of demonstrating a reasonable search for alternatives at the outset of a selection project.

After the fact, once adverse impact has been shown and defendants have adequately made the job-related argument, it is the plaintiffs' burden to demonstrate a feasible alternative to that job-related process that would have lesser adverse impact while retaining job-relatedness. One of the interesting aspects of this requirement is that plaintiffs may be able to meet their burden by demonstrating that there is an alternative method of using the selection procedure (e.g., an alternative method of weighting predictor components) that would have less adverse impact without affecting validity (Section 5[G] of the UGESP).

The evidence of the validity and utility of a selection procedure should support the method the user chooses for operational use of the procedure, if that method of use has a greater adverse impact than another method of use.

This principle was established in *Albemarle Paper Co. v. Moody* (1975). The Supreme Court ruled that if job-relatedness is proven, the plaintiff may prove pretext by showing that "other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship'." This principle was subsequently written into the UGESP and later codified in CRA-91.

Exemplar Case Law Related to Principle 3

Over the years, there have been unsuccessful attempts to prove job-related alternatives with lesser adverse impact (e.g., *Bridgeport Guardians, Inc. v. City of Bridgeport*, 1991). However, there were three 2006 district court rulings in which the principle was generally supported. These rulings are discussed in detail by Outtz (2007) and included *Ricci v. Destefano* (2006), *Bradley v. City of Lynn* (2006), and *Johnson v. City of Memphis* (2006). The *Ricci* ruling was overturned by the Supreme Court (*Ricci v. Destefano*, 2009), and the Johnson ruling was overturned by the 6th Circuit Court (*Johnson v. City of Memphis*, 2014). On the other hand, Bradley has not been overturned. In addition, one of the plaintiffs in *Ricci* (Michael Briscoe) sued, claiming that a 60% to 40% split between a written and oral exam produced adverse impact on minority candidates (*Briscoe v. New Haven*, 2011). We summarize these cases as follows.

The one case that has not been appealed is *Bradley v. City of Lynn* (2006), in which a cognitive test was the sole basis for selecting entry-level firefighters. Adverse impact was demonstrated, and the judge ruled there was insufficient evidence of job-relatedness. The judge also ruled that there were likely several equally valid alternatives with lesser impact, most notably the combination of cognitive and physical abilities (e.g., *Brunet v. City of Columbus*, 1995), as well as personality (work style) and biodata devices, which when combined with cognitive tests may produce less adverse impact than cognitive tests alone. The district court judge ruled: "while none of these approaches alone provides the silver bullet, these other noncognitive tests operate to reduce the disparate impact of the written cognitive examination."

In *Johnson v. City of Memphis* (2006), the judge ruled that an examination used for promotions to police sergeant was job-related, but ruled for the plaintiffs because of the existence of valid alternatives with lesser impact. Critical to this case was that the city had previously used a valid promotion test in 1996 and deviated from its prior procedure in the challenged test. The district court judge ruled: "It is of considerable significance that the City had achieved a successful promotional program in 1996 and yet failed to build upon that success." However, the ruling was reversed on appeal, where the 6th Circuit accepted the defendant's argument that the "protracted nature of simulation testing and the number of moving parts reinforced the City's

concerns about testing security." The 6th Circuit also accepted the defendant's argument that the video component took too long to administer and was too costly.

In *Ricci v. Destefano* (2006), the New Haven Civil Service Board (CSB) refused to certify promotional tests for lieutenant and captain positions because the Board had a "good faith" belief it would lose on adverse impact and offered an alternative (i.e., assessment centers) it believed was more valid with less adverse impact. However, the Supreme Court favored a higher standard—that the CSB needed "strong basis in evidence" for believing they would lose on adverse impact.

In *Briscoe v. City of Newhaven* (2010), Briscoe was the top scorer among 77 applicants for lieutenant on the oral exam, but he ranked 24th overall because of his poor performance on the written test. He challenged the 60–40 weighting favoring written tests on the basis that it was arbitrary and that a 70–30 weighting favoring the oral exam was as valid and would have less adverse impact on minorities. The district court judge rejected the claim based on the defense argument that the Board had a strong basis for believing it would lose on disparate treatment (an argument they never previously made). However, the Circuit Court overturned, rejecting this logic and remanded the case back for trial. This trial would have been interesting, but we will never know the outcome because Briscoe settled for $285,000 and a transfer to a director position.

So where does that leave the alternatives argument? The *Bradley* ruling still holds (so far), and the Briscoe settlement leaves room for the composite approaches advocated by De Corte (1999) and Sackett and Ellingson (1997). However, the reversal of the Johnson ruling is critical because, at least so far, it was the only ruling favoring alternatives in a case where the defendant had already won on proof of job-relatedness. So, we leave you with the same caveat as in the last edition—stay abreast of future decisions in this arena.

4. Work analysis should typically precede any selection process.

This is one point about which there is little disagreement. The SIOP *Principles*, the APA *Standards*, and the UGESP all clearly indicate that a demonstration of the validity of a selection device usually begins with a thorough analysis of the job if the criterion of interest is job performance. Nevertheless, if the criterion of interest is turnover, absenteeism, or other inherently important work outcomes, a full work analysis may not be necessary.

Work analysis generally refers to a detailed examination of a job, which may include the determination of what is done, how it is done, the context in which it is done (including the strategic importance of that job and its essential functions to organizational success), as well as the knowledge, skills, abilities, and other characteristics (KSAOs) required to perform the job successfully. The most common "O" characteristic is some measure of personality. Work analysis is also potentially important in demonstrating criterion-related validity because it establishes the importance or relevance of the criterion measure when some type of performance rating tool is used. Finally, work analysis is critical for demonstrating content validity because it provides a comparator for test content and establishes fidelity of the test content to the job content.

Proper work analysis is a critical requirement to meet professional standards for content-oriented validation. Work analysis should establish the linkage between important functions/characteristics of the job and KSAOs, and it should form the basis for linking a selection procedure to those KSAOs (or duties if the procedure is a work sample or simulation), thus linking the selection device to the job. The importance of work analysis to validation, particularly content validation, is clearly established in the UGESP, which provide the following guidance regarding the importance of work analysis:

> Validity studies should be based on review of information about the job. Any validity study should be based upon a review of information about the job for which the selection procedure is to be used. The review should include a work analysis. Section 14 (C2) is also instructive.
>
> (UGESP, Section 14(A))

Work analysis is also essential for content validity:

> There should be a work analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any work analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the work analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.
>
> (Section 1607.14(C)(2))

Another point to note is that work analysis is often critical to the definition of essential job functions in the Americans with Disabilities Act (ADA), as employers who exclude individuals for failure to perform nonessential job functions are in violation of this statute.

A final consideration is the frequency with which a work analysis should be updated. Although there is no specific guidance from regulatory agencies, many practitioners refresh work analyses every five to eight years regardless of significant changes to the job to ensure currency. Our view is that, all else being equal, five to eight years is no worse than any other rule of thumb, but the probative issue is whether the job has changed in meaningful ways.

Exemplar Case Law Related to Principle 4

The role of work analysis in criterion validity was established in *Griggs v. Duke Power* (1971) and *Albemarle Paper Co. v. Moody* (1975). Duke Power attempted to defend cognitive tests on the grounds that they were "professionally developed," yet they conducted no validity study. On the basis of the 1966 EEOC *Guidelines*, the Supreme Court ruled that professionally developed tests must:

> [F]airly measures the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant's ability to perform a particular job or class of jobs.
>
> (*Griggs v. Duke Power*, 1971, footnote 9)

The Albemarle Paper Company attempted to defend cognitive tests with a deficient criterion validity study conducted one month prior to trial. The most critical deficiency was the absence of work analysis. In *Moody*, the 4th Circuit ruled:

> In developing criteria of job performance by which to ascertain the validity of its tests, Albemarle failed to engage in any work analysis. Instead, test results were compared with possibly subjective ratings of supervisors who were given a vague standard by which to judge job performance. Other courts have expressed skepticism about the value of such ill-defined supervisor appraisals.
>
> (*Moody v. Albemarle*, 1973)

The Supreme Court affirmed the 4th Circuit ruling, and the Griggs and Albemarle rulings formed the basis of the 1978 UGESP.

The role of work analysis in content validity was established in *Guardians v. Civil Service* (1980) and has since been followed by every circuit that has ruled on content validity. Relevant cases for content validity are discussed under Principle 5. On the basis of *Guardians* and other cases, it is axiomatic that content validity cannot be established absent work analysis.

Finally, the importance of work analysis in the ADA is illustrated in *PGA v. Martin* (2001). Casey Martin, a professional golfer, who had a degenerative nerve disease and could not walk a golf course without pain, requested use of a golf cart during PGA tournaments as a reasonable accommodation. The PGA refused on grounds that walking the course is fundamental to PGA events. However, the Supreme Court unanimously ruled that walking is not "an essential attribute of the game." Similar examples include *Borkowski v. Valley Central* (1995) (whether

requirement to control students is an essential function for a librarian requesting a teacher's aide for that purpose) and *Stone v. City of Mt. Vernon* (1997) (whether a paraplegic former firefighter requesting a desk job must also fight fires). The point here is that essential functions must stand up to work analysis; otherwise, employers could easily define themselves out of the reach of the ADA without actual evidence.

5. Validation evidence for standardized tests and similar procedures can come in many forms and is not limited to one particular approach. In general, the greater number of converging sources of validation evidence, the better.

In the period immediately following passage of the Civil Rights Act of 1964, criterion-related validity was generally considered the gold standard for standardized tests, mainly because the 1970 EEOC *Guidelines* required that employment tests be "predictive of or significantly correlated with important elements of work behavior." Subsequently, the 1978 UGESP added content and construct-related validation to the arsenal of tools for demonstrating job-relatedness, thus finally cementing the "Trinitarian" view of validation, which had been introduced in the psychometric literature the 1950s. However, the UGESP contains the following warning on "appropriateness of content validity studies":

> A selection procedure based on inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures that purport to measure traits or constructs such as intelligence, aptitude, personality, common sense, judgment, leadership, and spatial ability.
>
> (UGESP, 1978, Sec. 1607.C (1))

For a short time, this passage perpetuated the myth that criterion-related validity is superior to content-related validity. By the mid-1980s, there was a growing consensus that discussions of "acceptable" models for validation were inappropriate because virtually any validation evidence, regardless of how gathered, is possible evidence of job-relatedness on its merits rather than by its name. This "Unitarian" view is now widely accepted by I-O psychologists and the courts. Furthermore, in recent decades, there have been several new and well-accepted techniques within and beyond the "Trinity," including the following as described by the SIOP Principles (2003):

- *Validity transport*: the use of a specific selection procedure in a new situation based on results of a validation research study conducted elsewhere[12]
- *Synthetic validity*: justification of the use of a selection procedure based upon the demonstrated validity of inferences from scores on the selection procedure with respect to one or more domains of work (job components)
- *Meta-analysis*: the accumulation of findings from a number of validity studies to determine the best estimates of the predictor-criterion relationship for the kinds of work domains

Therefore, when evaluating job-relatedness of standardized tests and similar procedures, it is often valuable to collect evidence from as many different sources or validation designs as feasible, without considering one particular design to be the best or only design. In general, the more evidence that is collected, the greater will be one's confidence in asserting job-relatedness.

Exemplar Case Law Related to Principle 5

The 1966 EEOC *Guidelines* were supported by the Supreme Court in *Griggs v. Duke Power* (1971) and *Albemarle Paper Co. v. Moody* (1975). However, after adopting the UGESP in 1978, courts immediately essentially overruled UGESP by supporting content validity for inferences about mental processes. The landmark case supporting content validity was the 2nd Circuit ruling in *Guardians v. Civil Service* (1980), which outlined five steps for content validity:

1. Suitable work analysis
2. Reasonable competence in test construction
3. Test content related to job content
4. Test content representative of job content
5. Scoring systems selecting applicants who are likely to be better job performers

In *Guardians v. Civil Service* (1980), although the defendants could not support a cut score or rank ordering on the basis of weakness in steps 2 and 3 (see Principle 8 below), they were able to use a content validation strategy for demonstrating job-relatedness. In *Gillespie v. State of Wisconsin* (1985), the 7th Circuit ruled "neither the *Uniform Guidelines* nor the psychological literature express a blanket preference for criterion-related validity" and in *Police Officers v. City of Columbus* (1990), the 6th Circuit, citing the 1987 SIOP *Principles*, ruled that it is critical that "selection instruments measure a substantial and important part of the job reliably, and provide adequate discrimination in the score ranges involved." Content validity was then supported in many subsequent cases, including by the 2nd Circuit in *Gulino v. New York State Education Department* (2006).

Since the prior edition of the Handbook, there have been several additional confirmations of the sufficiency of content evidence, including (1) a more recent ruling in *Gulino v. New York State Education Department* (2012), in which the content validity of a teacher licensing examination was struck down for failure to meet all five *Guardian* criteria, and (2) *Smith v. City of Boston* (2015), in which the content validity of a police promotion exam to sergeant was struck down because of failures in steps 4 (test content representative of job content) and 5 (reliable scoring system) from *Guardians*. However, in *Lopez v. City of Lawrence* (2014), the content validity of police promotion exams to sergeant was upheld, even though the judge accepted the opinion of one of the experts in that case that the exams were "minimally valid."

On the other hand, attempts to support job-relatedness on the basis of meta-analysis alone have often failed (*EEOC v. Atlas Paper*, 1989; *Lewis v. Chicago*, 2005). Nevertheless, meta-analysis has been credited as supplemental to local validity studies (*Adams v. City of Chicago*, 1996; *Williams v. Ford Motor Co.*, 1999). Also, transportability and corrections for statistical artifacts were supported in *Bernard v. Gulf Oil Corp.* (1989), in which criterion validity was found for two of five jobs, and the 5th Circuit found "sufficient similarity in the skills required" for all five jobs. The Court also ruled "the adjusted figures . . . are better estimates of validity," and uncorrected coefficients "underestimate" validity of the tests.

In summary, although any of the "Trinity" approaches alone are likely sufficient for job-relatedness, a notion consistent with the UGESP, the body of case law as a whole suggests that a "Unitarian" approach leveraging multiple sources of validation may have greater acceptance in court and that new methods of supporting validity are at least acceptable supplements to traditional methods.

A final point to note is that each of us has been involved in content validity cases, some still ongoing as this principle was being written, and it is our opinion that step 2 in *Guardians v. Civil Service* (1980) (Reasonable competence in test construction) often requires greater expertise for tests involving simulations than for multiple-choice tests. Therefore, our advice to those entities choosing to use simulations is to consider enlisting outside expertise from among those with a proven track record in designing such exams.

6. In a selection context, criterion information, particularly in the form of performance ratings, is as important in validation as is predictor information.

For more than 50 years, I-O psychology has recognized the importance of the criterion in a human resource (HR) system. In the context of validation, criterion information also represents the second "half" of the criterion-related validation model. Criterion information appears often in the form of supervisor ratings of performance. As is the case with predictors, the credibility of criterion information should be established through job analysis information, supported via psychometric research (e.g., demonstrating reliability or factor structure), or both when possible.

Performance evaluations often play a significant role in non-entry-selection decisions such as promotions, training assignments, job changes, and reductions in force. To the extent that criterion information becomes predictor information (e.g., performance ratings are at least a partial foundation for a personnel decision such as promotion or downsizing), then that criterion information may be considered part of the selection process and analyzed in a way that permits the inference that this performance information was job-related, psychometrically credible, and fair. Examinations of criterion information often involve not only substance issues but also process issues such as the right of an employee to "appeal" information, as well as the extent to which those providing criterion information are knowledgeable and competent judges of the employee's performance.

Exemplar Case Law Related to Principle 6

Performance appraisal was a key feature in two landmark Supreme Court rulings on adverse impact: *Albemarle Paper Co. v. Moody* (1975), a hiring case, and *Watson v. Fort Worth Bank* (1988), a promotion case. It was also a key feature in *Meacham v. Knolls* (2006).

There were multiple deficiencies in the criterion validity study conducted by the defendant in *Albemarle*. Chief among them was the failure to establish a reliable and valid criterion against which test scores were compared. In the words of the Supreme Court, "The study compared test scores with subjective supervisorial rankings." Although UGESP allow the use of supervisorial rankings in test validation, they quite plainly contemplate that the rankings will be elicited with far more care than was demonstrated here. Albemarle's supervisors were asked to rank employees by a "standard" that was extremely vague and fatally open to divergent interpretations. Each "job grouping" contained several different jobs, and the supervisors were asked, in each grouping, to "determine which ones [employees] they felt, irrespective of the job that they were actually doing, but in their respective jobs, did a better job than the person they were rating against."

In *Watson*, the main challenge was subjective ratings of job performance. Clara Watson also challenged subjective ratings of interview performance and past experience. The main issue related to more subjective procedures that could cause adverse impact. The Supreme Court unanimously ruled there can be subjective causes of adverse impact in an 8–0 decision.

In *Meacham*, 30 of 31 employees laid off in an involuntary reduction in force (IRIF) were over age 40. The layoffs were based entirely on performance appraisal ratings. In the words of the 2nd Circuit, the key RIF criteria were "subjective assessments of criticality and flexibility" of employee skills (*Meacham v. Knolls*, 2006).

In summary, irrespective of whether the issue is hiring, promotion, termination, or the criterion in a criterion validity study, the *Albemarle*, *Watson*, and *Meacham* rulings carry inherent warnings that performance appraisals should be based on work analysis and that the methodology used to appraise worker performance meet acceptable psychometric properties.

7. The evidence required to demonstrate job-relatedness in the legal context typically differs for biographical factors (e.g., educational requirements) and physical factors (e.g., height and weight) than it does for standardized tests.

Principle 5 discusses approaches to demonstrating job-relatedness for standardized tests, but there are sources of adverse impact other than standardized tests, and these can serve as MQs for selection, particularly in hiring. As discussed in Principle 5, demonstration of job-relatedness for standardized tests is exacting. In comparison, case law reveals that legal proof of job-relatedness for biographical MQs is often less exacting and legal proof of job-relatedness for physical factors is often more exacting than legal proof of job-relatedness for standardized tests. In fact, Gutman et. al (2010) refer to defending biographical MQs as "adverse impact light" and defending physical factors as "adverse impact heavy." In this scheme, then, defending standardized tests falls somewhere in between (call it "adverse impact moderate").

Exemplar Case Law Related to Principle 7

The term "adverse impact light" implies it is generally easier to prove job-relatedness with biographical variables, particularly if public safety is threatened. The two most extreme examples of the light defense are *Hyland v. Fukada* (1978) and *NYC v. Beazer* (1979). In Fukada, the 9th Circuit accepted articulated safety concerns to exclude from a security guard position a felon who had been previously convicted of armed robbery from a security job. In *Beazer* (1979), the Supreme Court upheld exclusion of methadone users for the position of transit authority cop because it is obvious that drug addiction threatens the "legitimate employment goals of safety and efficiency." Similarly, in *Davis v. Dallas* (1985), the 5th Circuit accepted an articulated reason for excluding recent drug users from police work—that it shows a disregard for the law.

In some cases, stronger evidence has been required in cases involving biographical variables, but none rising to the defense for standardized tests in the UGESP. For example, in *Spurlock v. United Airlines* (1972), a black applicant was excluded from flight officer training because (1) he had only 204 hours of flight time (500 were required) and (2) he had only two years of college (a four-year degree was required). The defense for the flight time variable showed a significant negative correlation between flight hours and training failures. However, on the degree requirement, an expert testified that a four-year degree was necessary to "cope" with rigorous classroom training requirements.

Other examples of the lighter defense include *United States v. Buffalo* (1978), where a high school diploma for police officers was upheld based on federal commission reports in the 1960s that "a high school education is a bare minimum requirement for successful performance of the policeman's responsibilities." Also, in *Davis v. Dallas* (1985), the 5th Circuit upheld a requirement of 45 hours of college credit with C or better grades for police officers based on the task force reports cited in *United States v. Buffalo* (1978), as was a "poor driving" exclusion based on research indicating that past driving habits predict future driving habits.

The adverse impact "heavy" designation implies a greater defense burden for physical requirements as compared to standard tests. The best illustration of this defense is *Dothard v. Rawlinson* (1977), where the Supreme Court rejected a height and weight criterion that adversely impacted women applying for prison guard positions in an all-male maximum security prison. After failing in defense of this requirement, the State of Alabama argued successfully that it is necessary to exclude all women from the job because 20% of the prison population included sex offenders, and having women in this situation presented an extra threat to prison safety. Thus, it was easier to defend exclusion of all women than it was the exclusion of all people who did not meet minimal height and weight requirements. Stated differently, exclusion based on pure physical requirements, in effect, rose to the level of the Bona Fide Occupational Defense (BFOQ).[13]

In other cases, employers have sometimes succeeded in defending physical requirements, as in *Boyd v. Ozark Air Lines* (1977), where the airline proved that shorter pilots could not safely operate all cockpit instruments. However, the more typical result in such cases is illustrated in *Horace v. Pontiac* (1980), where the police department asserted that being tall is necessary for police officers to fend off and gain the respect of criminals, to which the 6th Circuit responded that there were superior and more direct methods of assessing such capabilities. Critically, in *Boyd v. Ozark Air Lines* (1977), the physical requirement was a critical attribute.

Additionally, consider *Bradley v. Pizzaco of Nebraska, Inc.* (1993), where the charge was adverse impact against blacks that occurred because of a "no beards" policy. In its defense, Domino's Pizza offered survey data indicating customer preference for cleanly shaven counter and delivery staff. However, the 8th Circuit struck down this defense, ruling "the existence of a beard on the face of a delivery man does not affect in any manner Domino's ability to make or deliver pizzas to their customers." In comparison, a "no beards" policy was upheld in *Fitzpatrick v. Atlanta* (1993), because the City of Atlanta proved it is essential for firefighters to be beardless for facial safety equipment to function properly, thus posing a danger to other firefighters, as well as civilians whose lives might be in danger.

In summary, especially when public safety is a primary concern, courts have a history of accepting less rigorous defenses for biographical variables than they have required for

standardized tests. On the other hand, the defense for pure physical characteristics boils down to proof that the requisite characteristic is necessary for the business entity to survive. This standard imposes a heavier defense burden than for standardized tests because evidence of (for example) content validity of a test does not imply that a business entity is terminally threatened if the test is not used.

8. Cut scores should be based on a rational foundation that may or may not include empirical analyses.

There are several types of cut scores. The first is a *nominal cut score*. This is a value often established as an arbitrary pass/fail score in a multiple-hurdle system. It designates the threshold for continuing to the next stage in the process. The second is known as the *effective cut score*. This is the score below which no one is hired or appointed. It is not predetermined but simply identified after the fact. Such scores are most often seen in strict rank-order appointments where candidates are appointed from the top scorer down until all positions are filled. The score of the individual filling the last opening is the effective cut score. The third type of cut score is the *critical cut score*. This value has been chosen to represent the score below which an applicant is thought to fall below a minimal standard for effective job performance. Often, a critical cut score is tied to the safety or well-being of the candidate or the public/customer base via some type of criterion-related analysis or subject matter expert judgement.

Nominal cut scores are often set by combining practical and theoretical issues. A practical consideration might be the cost of testing. Assume there are 1,000 applicants for 20 openings and there will be a multistage assessment process. The employer might want to limit the cost of testing by eliminating many individuals at the first stage of the process. In this case, the cut score can be set by looking at the selection ratio and deriving some estimate of acceptable performance expectations for candidates.

There is no need to "set" an effective cut score. It is an axiomatic score determined solely by the score of the last person hired or appointed in a rank-ordered list of candidates. With respect to a critical cut score, there are several options available, but all require some consideration of a criterion level that distinguishes between competent and incompetent or minimally qualified and less-than-minimally qualified. Subject matter experts can estimate requisite predictor performance and associated criterion performance. Such estimates, of course, would benefit greatly from a comprehensive and accurate work analysis. Incumbent populations can be used to identify predictor scores associated with minimally acceptable performance. If predictor and criterion data are available, analyses could be conducted to identify a critical cut score. Regardless of which techniques are used to set the critical cut score, there should be a rational foundation for its choice or an empirical one based on a work analysis. Although it is axiomatic that a single score (i.e., a cut score) cannot be "validated," it is possible to produce evidence that provides some confidence that a critical cut score is related to an anticipated real-world outcome.

Exemplar Case Law Related to Principle 8

Until the 3rd Circuit's ruling in *Lanning v. SEPTA* (1999), all courts relied on the UGESP, which state: "Where cut-off 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" (Section 1607.5(H)). For example, in *Guardians v. Civil Service* (1980) and *Gillespie v. State of Wisconsin* (1985), two cases discussed earlier under Principle 5, the 2nd and 7th Circuits ruled that:

> An employer may establish a justifiable reason for a cut-off score by, for example, using a professional estimate of the requisite ability levels, or, at the very least by analyzing the test results to locate a logical break-point in the distribution of scores.

Both studies employed content validity strategies. However, the defendants lost on cutoff score in *Guardians v. Civil Service* (1980) but won in *Gillespie*.

In *Guardians*, the defendants made selections on strict rank-ordering, and defined the cutoff score at below the point that the last applicant was selected (effective cut score). The *Guardians* Court ruled:

> If it had been shown that the exam measures ability with sufficient differentiating power to justify rank-ordering, it would have been valid to set the cutoff score at the point where rank-ordering filled the City's needs. . .. But the City can make no such claim, since it never established a valid basis for rank-ordering.

However, in *Gillespie*, the 7th Circuit ruled in favor of the defendant on the basis of two factors: (a) establishment of inter-rater reliability (of grades) and (b) the cutoff was selected to permit interviewing "as many minority candidates as possible while at the same time assuring that the candidates possessed the minimum skills necessary to perform" the job. On rank ordering, the 7th Circuit cited verbatim from Q62 of the "Questions and Answers" of the UGESP:

> Use of a selection procedure on a ranking basis may be supported by content validity if there is evidence from work analysis or other empirical data that what is measured by the selection procedure is associated with differences in levels of job performance.

Therefore, up until the *Lanning* ruling, there was little dispute among the circuit courts on either cutoff scores or rank ordering.

In *Lanning*, the 3rd Circuit interpreted the terms "job-related" and "consistent with business necessity" from the Civil Rights Act of 1991 as implying separate standards and that the "business necessity" part implied proof that the cutoff score "measures the minimum qualifications necessary for successful performance of the job in question." This interpretation was explicitly rejected in *Bew v. City of Chicago* (2001), where the 7th Circuit ruled: "Griggs does not distinguish business necessity and job relatedness as two separate standards."

All other circuits have continued to follow precedents from *Guardians* and *Gillespie*, with the exception of a district court ruling in *Isabel v. City of Memphis* (2005), which was affirmed by the 6th Circuit on appeal. However, there was no need for either court to cite *Lanning* because the city lost on other grounds. For example, the city's expert, who designed the test, admitted in open court that the job knowledge component did not represent the full job domain, and he testified that the cutoff point chosen was "totally inappropriate," a "logical absurdity," and "ludicrous."

In summary, in our opinion, the *Guardians* ruling, along with UGESP guidance cited above, remain the most important basis for establishing cutoff scores, particularly in content validity studies.

9. An optimal balance between job-relatedness and reduction of adverse impact should be struck when possible.

The key point here is the definition of *optimal*. Several researchers have demonstrated that optimum validity depends upon the manner in which job performance is defined or the specific aspect(s) of performance that are most important to an employer (Hattrup, Rock, & Scalia, 1997; Murphy & Shiarella, 1997). Murphy and Shiarella (1997) proposed that weighting of predictors and criteria provides a better understanding of the relationship between selection and job performance. They documented that the validity of a predictor composite can vary substantially depending upon the weight given to predictors and criterion measures. The 95% confidence interval for the validity coefficients for various weightings varied widely from as low as .20 to as high as .78.

Another challenging aspect of balancing job-relatedness and adverse impact is the fact that job-relatedness can be defined via different levels of analysis. When individual productivity and task performance are the focus, cognitive ability tests (instruments that typically have high

adverse impact) typically result in the highest validity for a single predictor. However, if overall organizational effectiveness is the objective, factors such as legal defensibility, strategic positioning within the marketplace, employee performance, workforce diversity, and corporate social responsibility may all be considered.

The bottom line is that the organization's mission and values dictate what constitutes acceptable performance and, ultimately, the best methods of achieving that performance. This argument was made quite forcefully in the University of Michigan Law School admission cases *Grutter v. Bollinger* (2003) and *Gratz v. Bollinger* (2003). The University of Michigan took the position that its objective was to admit a first-year class that collectively advanced the law school's overall mission as opposed to simply admitting each student with the highest probability of achieving a given law school grade point average (GPA). Whether one agrees with the school's stated mission or not, once formulated, that mission basically defines job-relatedness. It also drives the types of strategies that are most likely to achieve an optimum balance between job-relatedness and reduction of adverse impact.

Exemplar Case Law Related to Principle 9

Two methods of reducing adverse impact have found favor among the courts: (a) eliminating test battery components that are most likely to produce adverse impact and (b) using other factors (e.g., diversity) in the selection process that are more likely to benefit minorities. *Hayden v. County of Nassau* (1999) illustrates the first method and the cases *Grutter v. Bollinger* (2003) illustrate the second.

After losing in prior litigation, and facing a consent decree, Nassau County New York was motivated to develop a hiring exam for police officers that reduced or eliminated adverse impact. A 25-component test battery was initially administered to 25,000 candidates. The goal of reducing (but not eliminating) adverse impact was accomplished by eliminating scores on 16 of the 25 components. Another configuration with even less adverse impact (and even fewer components) was rejected because it had lower validity. This process was challenged by 68 unsuccessful nonminority candidates who would have benefited if all 25 components had been maintained. The 2nd Circuit favored Nassau County, ruling that "the intent to remedy the disparate impact of the prior exams is not equivalent to an intent to discriminate against non-minority applicants."

In *Grutter*, the Supreme Court ruled (under the Fourteenth Amendment) that (a) diversity is a compelling government interest and (b) the method used by the law school was narrowly tailored to that interest. At the same time, the Supreme Court struck down the Michigan undergraduate diversity plan for not being narrowly tailored (*Gratz v. Bollinger*, 2003). The *Grutter* ruling was based on Justice Powell's 1978 ruling in *Regents of University of California v. Bakke* (1978). Between *Bakke* and *Grutter*, several courts upheld preference for minorities based on diversity in police forces, including *Detroit Police Association v. Young* (1979) and *Talbert v. City of Richmond* (1981). After *Grutter*, the 7th Circuit upheld preference for Black police officers in a promotion process, ruling in *Petit v. City of Chicago*, 2003, noting:

> It seems to us that there is an even more compelling need for diversity in a large metropolitan police force charged with protecting a racially and ethnically divided major American city like Chicago. Under the *Grutter* standards, we hold, the city of Chicago has set out a compelling operational need for a diverse police department.

The 7th Circuit then upheld out-of-rank promotions of minority applicants on grounds that it was narrowly tailored.

It should be noted that the key ingredient in a successful defendants' diversity argument is often proving to the satisfaction of the courts that diversity is an important job-related factor that furthers the mission of the organization. Diversity for diversity's sake, therefore, will typically not work. For example, in *Lomack v. City of Newark* (2006), a newly elected mayor transferred firefighters to and from various posts so that all 108 fire stations were racially diverse. The mayor felt there were "educational and sociological" benefits for such a "rainbow," but the

3rd Circuit saw it as "outright racial balancing." Similar rulings were rendered in *Biondo v. City of Chicago* (2004) and *Rudin v. Lincoln Land Community College* (2005).

Two recent cases are also worth considering. First, *Parents v. Seattle School District* (2007) involved two school districts, each with plans to fight de facto segregation. One plan (Seattle) featured three tiebreakers for admission into any of 10 high schools, one of which was race-based and the other plan (Jefferson County) involved clusters of school with targets of 15% minimum and 50% maximum per cluster. Both plans were struck down. However, Justice Kennedy ruled for a majority of five that race is still a compelling state interest. Although he also ruled the plans were not narrowly tailored, he offered solutions he felt were race-neutral and narrowly tailored.

Second, just before this chapter was submitted, the Supreme Court ruled that the University of Texas's admissions policy that considered diversity was legal, thus affirming the *Grutter* ruling. The case is *Fisher v University of Texas at Austin (2013)* and the issues in the matter are complex.[14] Basically, the University of Texas found it was lacking in the "critical mass" of minority students, and in part in response to the *Grutter* ruling, developed a plan that permitted racial preference for a small percentage of students.

10. Many of the same practices that define responsible selection define responsible downsizing efforts.

Downsizing (also known as reductions-in-force, or RIFs) represents a special instance of selection. Individuals are typically selected to "stay" in an organization (or conversely to "leave" an organization). It might be thought of as "deselection." It represents the mirror image of the selection scenario. However, there are some unique aspects to downsizing. In the context of litigation, the employer typically must be prepared to show that the RIF was not a pretext for simply eliminating members of a protected group (e.g., female, minority, and/or older employees). Thus, the RIF may be tied to a larger business plan that documents the need for the reduction, which may also support why certain jobs, departments, or divisions have been targeted for the force reduction.

As is the case in many selection/promotion scenarios, the employer should typically identify the particular knowledge, skills, abilities, or other personal characteristics that are central to the decision about whom to lay off. These variables might include abilities and skills needed for future vitality of the organization, critical experience bases, customer contacts, past performance, and future output responsibilities. Many of these variables may be illuminated by a current or future-oriented work analysis. In addition, selection procedures used to evaluate current employees (e.g., performance ratings, skill ratings, knowledge ratings) should typically conform to accepted professional and scientific standards for the use of such rating devices.

Unlike selection scenarios, it is often difficult to formally "validate" a downsizing process because there is no obvious criterion for success beyond simple organizational survival. Nevertheless, just as in the case of selection, it is important to develop a theory of the downsizing process and its desired results. This theory would typically include some statement of requisite KSAOs for going forward, as well as the organizational need for downsizing.

Exemplar Case Law Related to Principle 10

Most RIF cases are age-based and involve disparate treatment charges. However, some are race-based (e.g., *Jackson v. FedEx*, 2008). Furthermore, after the Supreme Court's ruling in *Smith v. City of Jackson* (2005), which clarified that adverse impact is a valid ADEA claim, we can expect more age-based adverse impact cases even in RIFs (e.g., *Meacham v. Knoll*, 2006). In a meta-analysis of 115 district court cases involving disparate treatment based on age, Wingate, Thornton, McIntyre, and Frame (2003) found that 73% of the rulings were summary judgment for defendants (SJDs). Factors associated with SJD were use of (a) performance appraisal, (b) organizational review, (c) employee assessment and selection methods, and (d) a concrete layoff plan. As a general principle, employers establishing and following sound concrete layoff policies will likely prevail.

Cases with favorable rulings for plaintiffs reveal possible key mistakes made by employers. The most obvious mistake is weakness in the layoff plan. For example, in *Zuniga v. Boeing Company* (2005), the defendant had a concrete layoff plan that relied heavily on performance evaluations. However, the plaintiff defeated SJD by proving that the evaluations he received during the layoff process were inconsistent with performance evaluations he received shortly before the plan was established.

Employers have made mistakes in reassignment after the RIF. For example, in *Berndt v. Kaiser Aluminum & Chemical Sales, Inc.* (1986), Berndt could not compete for other jobs in which he was arguably more qualified than younger employees who were afforded this opportunity. Similarly, in *Zaccagnini v. Chas. Levy Circulating Co.* (2003), older truck drivers were not considered for newly available jobs that younger drivers received.

There are also cases where reassignment is part of the RIF plan, but the definition of "similarly situated" older versus younger employees is narrowly construed. For example, in *Ercegovich v. Goodyear Tire and Rubber Co.* (1998), three HR positions were eliminated, and the oldest employee was not reassigned. The defendant argued that the three employees were not similarly situated because they performed different job functions, but the 6th Circuit ruled for the plaintiff because the three eliminated jobs involved common knowledge, skills, and abilities.

Plaintiffs have also successfully used direct evidence of stray remarks by supervisors. For example, in *Starceski v. Westinghouse Elec. Corp.* (1995), a supervisor admitted he stated that "it was actually a fact that older engineers . . . were going to be let go" (see also *Madel v. FCI Marketing*, 1997).

Finally, employers must strictly follow eight explicit requirements in the Older Workers Benefit Protection Act of 1990 if they offer enhanced benefits to older employees who accept early retirement in exchange for waiver of the right to sue. These requirements are that (a) the waiver document is clearly written and easily understood; (b) the waiver document cites the ADEA; (c) it only affects rights prior to the effective date of the waiver; (d) it offers enhanced benefits; (e) it advises employees of their right to seek counsel; (f) it provides 21 days for individuals and 45 days for groups to make a decision; (g) it is revocable within 7 days of signing; and (h) it provides extensive information about who is affected if a group is involved. In *Oubre v. Entergy* (1998), the Supreme Court made it clear it will strictly construe these eight requirements.

11. Employers should typically take proactive measures to prevent, detect, and correct EEO violations involving procedural unfairness to workers or violations inferred from analysis of workforce data.

A potentially effective model for preventing, detecting, and correcting EEO violations involving unfair treatment of employees is provided in exemplar form by EEOC's approach to sexual harassment (SH). In EEOC Policy Guidance 915.050 (June 1999; http://www.eeoc.gov/policy/docs/currentissues.html), the EEOC distinguished between minimum requirements and best practices. The minimum requirements are as follows:

- A clear explanation of prohibited conduct;
- Assurance that employees who make complaints of harassment or provide information related to such complaints will be protected against retaliation;
- A clearly described complaint process that provides accessible avenues of complaint;
- Assurance that the employer will protect the confidentiality of harassment complaints to the extent possible;
- A complaint process that provides a prompt, thorough, and impartial investigation;
- Assurance that the employer will take immediate and appropriate corrective action when it determines that harassment has occurred.

An aggressive response requires additional actions, including training employees to understand all employer policies, a dedicated "EEO Officer" to handle complaints, and an employee handbook that summarizes the rights and privileges for all employees.

Although Policy Guidance 915.050 was expressly targeted at SH, it applies equally well to broader actions of any sort. For example, employees may feel that their performance is being improperly appraised or that they are not receiving training or certification opportunities necessary for advancement.

Workforce data may include confidential survey information obtained by a trained EEO Officer or statistical data relating to performance of different groups of employees on selection tests, composition of the workforce in relation to the appropriate labor pool, or the funneling of different groups into different jobs (e.g., females offered jobs as cashiers vs. males offered jobs as assistant managers). Properly analyzed data could lead to the detection of a potential adverse impact or pattern and practice violation, thus enabling employers to take action before expensive and disruptive litigation while simultaneously increasing employee perceptions of fairness (see, for example, McPhail, which details issues related to "self-critical analyses" that could be used against employers).[15]

Finally, employers need a good anti-retaliation policy. EEOC statistics reveal that retaliation complaints are increasing, although EEO claims in general have stabilized and even decreased (Zink & Gutman, 2005). Adding to this caution, the Supreme Court recently lightened the burden on plaintiffs to prove retaliation, requiring employers to educate their HR professionals and managers on what to do when individuals complain about a workplace policy or file formal charges. It is not unusual for plaintiff employees to complain about routine selection decisions, such as lost training opportunities that might presage promotions, or failure to promote per se. It is critically important for employers to prevent reprisals against those who do complain.

Exemplar Case Law Related to Principle 11

EEOC Policy Guidance 915.050 interprets the Supreme Court's 1998 rulings in *Burlington Industries, Inc. v. Ellerth* (1998) and *Faragher v. City of Boca Raton* (1998). The Court ruled in both cases that even when "no tangible employment action is taken" the employer has vicarious liability for supervisors, but may affirmatively defend itself by proving with evidence that it exercised (a) "reasonable care to prevent and correct promptly, sexually harassing behavior" and (b) "the . . . employee unreasonably failed to take advantage of any preventive or corrective opportunities provided by the employer to avoid harm otherwise."

Ellerth applies to private entities and *Faragher* to public entities. Both rulings clarify that (a) there is no defense for *quid pro quo* SH (strict liability); (b) there is an affirmative (see above) defense for hostile harassment by supervisors; and (c) employers must know or have a basis for knowing that SH occurred among coworkers or they will be considered guilty of reckless disregard. Examples of employer policies that succeeded in affirmative defenses are in *Coates v. Sundor Brands, Inc.* (1998) and *Shaw v. AutoZone* (1999), and examples of employer failures affirmatively to defend are in *Baty v. Willamette Industries, Inc.* (1999), *Dees v. Johnson Controls* (1999), and *Gentry v. Export Packing* (2001).

There are two major Supreme Court rulings on retaliation. In *Robinson v. Shell Oil* (1997), the Court unanimously ruled that retaliation applies to actions of a former employer who wrote a negative letter of reference for a previously fired employee. Subsequently, the EEOC issued policy guidance (915.003) in May 1998 outlining three steps for proving retaliation: (1) opposing an employer policy (opposition) or filing a legal claim (participation), (2) suffering an adverse action, and (3) causally connecting opposition or participation to the adverse action. At the same time, the EEOC defined "adverse action" as any action reasonably likely to deter charging parties (or others) from engaging in a protected activity.

More recently, in *Burlington N. & SFR Co. v. White* (2006), the Supreme Court endorsed the EEOC's definition of "adverse action" over two earlier but heavier standards. One of those heavier standards required "adverse action" to include ultimate employment consequences such as hiring, discharge, promotion, or compensation. The other required proof of interference with terms and conditions of employment. By endorsing a lighter EEOC standard, the Supreme Court made it possible for otherwise legal employer actions

to constitute retaliation. For example, in *Moore v. Philadelphia* (2006), White police officers opposed harassment against fellow Black police officers, and in *Hare v. Potter* (2007), a female employee cited 10 incidents in which her life was made miserable after she filed an EEOC complaint. These actions were deemed insufficient for proof of both racial and sexual harassment but were deemed sufficient to prove retaliation against the complainants, although they were not original plaintiffs.

Finally, the risks employers face when altering selection processes to avoid adverse impact are illustrated in *Ricci v. Destefano* (2008) and *Hayden v. County of Nassau* (1999). In *Ricci*, where exams were discarded, there was a weak basis for the New Haven Civil Service Board (CSB) to believe they would lose an adverse impact challenge to minority applicants. However, in *Hayden*, where Nassau County (New York) eliminated portions of the test to reduce the adverse impact, there was a much stronger basis for that fear since the county was under a consent decree to create a valid test with the least amount of adverse impact.

12. Adverse impact is a valid claim in both Title VII and the ADEA. However, there are critical differences in each of the three phases in how the Title VII and ADEA scenarios are tried in court.

There are three phases in an EEO trial. Phase 1 consists of prima facie evidence of a "substantial difference" in selection rates between two protected groups, Phase 2 is a defense against the prima facie claim, and Phase 3 is proof that the defense in Phase 2 is a pretext for discrimination.

In the Title VII scenario, most Phase 1 claims involve statistical evidence that a test or other selection procedure produces "substantial differences"[16] in selection rates between two groups (e.g., Blacks scoring lower than Whites on IQ tests in *Griggs v. Duke Power*, 1971). However, it is also possible to show adverse impact with minimal qualifications (MQs) that are chilling factors (e.g., having a high school degree; also in *Griggs v. Duke Power*). Either way, if adverse impact is shown, the defendant must prove in Phase 2 that what caused the adverse impact is job related and consistent with business necessity, forcing the plaintiff to prove in Phase 3 that alternative selection procedures are equally valid and produce less or no adverse impact in comparison to the test or selection procedure challenged. As noted above, each of the three phases in the ADEA differ from the Title VII prescription in part or in whole.

The Phase 1 difference is partial; differences between older and younger workers due to a selection procedure (e.g., reduction in force, or RIF) are treated in the same way as racial differences based on a selection procedure. What is different is that in Title VII a correlation between (for example) an MQ and sex (e.g., minimum height/weight criteria that exclude more women than men) is sufficient to make the *prima facie* case, whereas a correlation between an MQ and age (e.g., higher percentage wage increases for individuals with fewer years of service) is not sufficient for the *prima facie* case in the ADEA because decisions correlated with age are not necessarily motivated by age.

Second, as noted in Principle 5, there are three different types of Phase 2 defenses in Title VII (i.e., for standardized tests, biographical factors, and physical factors), each of which is designed to show that a cause of adverse impact is job related and consistent with business necessity. The ADEA uses the statutory Reasonable Factors Other Than Age (RFOA) defense. This defense is entirely different from Title VII. Additionally, the reasonable factor(s) must be proven with evidence, and ***not*** with a simple articulation as in most disparate treatment cases.

Third, the proof of pretext in Phase 3 is entirely different in the ADEA compared to Title VII. As noted in Principle 3, the Title VII pretext phase requires proof of equally valid alternatives that result in less or no adverse impact. The parallel argument (i.e., a "more reasonable" factor other than age) is not available in the ADEA. Rather, the requirement here is much the same as in typical disparate treatment cases—that the factors offered in Phase 2 are not the true factors but rather a cover-up for discrimination.

Exemplar Case Law Related to Principle 12

In the 1980s, courts treated adverse impact in age cases with Title VII rules. For example, cost-cutting defenses in *Geller v. Markham* (1980) (hiring at the lowest of six steps) and *Leftwich v. Harris Stowe State College* (1983) (termination of tenured faculty) failed because neither was deemed job related in accordance with then existing DoL regulations (subsequently adopted by the EEOC). Under current rules, these factors are correlated with age and not necessarily motivated by age.

The motivation requirement was introduced in *Hazen v. Biggins* (1993), where a 62-year-old was terminated shortly before eligibility for pension vestment, a clear-cut ERISA violation. However, the lower courts also favored disparate treatment because age and years of service are correlated. The Supreme Court reversed on disparate treatment, ruling unanimously that employer decisions may be motivated by "factors other than age . . . even if the motivating factor is correlated with age." Additionally, three justices opined that it is "improper to carry over disparate impact analysis from Title VII to the ADEA." After *Hazen*, three circuit courts continued to entertain age-based adverse impact claims, but seven circuit courts found adverse impact inapplicable in the ADEA as a matter of law.

Then, in *Smith v. City of Jackson* (2005), police officers and dispatchers with less than five years of experience received higher percentage compensation increases. The lower courts ruled adverse impact was unavailable in the ADEA, but the Supreme Court ruled that Hazen does not preclude such claims. However, the Supreme Court affirmed in *Hazen* that factors correlated with age (e.g., years of service) do not qualify, and where adverse impact is shown, defendants may use the statutory RFOA in lieu of proving job-relatedness. The plaintiffs ultimately lost the *prima facie* case (failure to show adverse impact), and the City had a valid RFOA (the need to compete with neighboring municipalities for filling lower-level positions by increasing the compensation for those positions, although the entry-level positions were often filled by younger applicants).

Although it solidified adverse impact as a valid ADEA claim, the *Smith* (2005) ruling was not clear on whether the RFOA required an articulation or actual proof. This ambiguity is illustrated in *Meacham v. Knolls Atomic Power Lab (KAPL)*, which was reviewed by the 2nd Circuit Court both before (*Meacham* I, 2004) and after (*Meacham* II, 2006) the *Smith* ruling. In this case, 30 of 31 employees laid off in an RIF were over age 40. Using pre-*Smith* rules, the court found there were alternatives with less adverse impact in Meacham I. However, in Meacham II, after the *Smith* ruling, the court found that KAPL had *articulated* two nondiscriminatory reasons—whether employees were "flexible" and "retrainable" for alternative assignments. Recognizing, perhaps, the confusion created in the *Smith* ruling, the Supreme Court ruled in *Meacham v. KAPL* (2008) that the RFOA defense is affirmative, and requires evidence, not merely articulation.

As a postscript, it is not clear what actual proof KAPL needed to provide. When the case was ultimately returned to the district court (*Meacham v. KAPL*, 2009), the court ruled that the defendants had waived the right to make the RFOA proof. Nevertheless, the moral of the story is to take a conservative route in conducting an RIF, such that specific measurable selection factors and evidence of their reliability are essential parts of the layoff plan.

13. All disability-related decisions should be made on a case-by-case basis, including determining if an applicant or employee is (a) disabled, (b) needs accommodations for performing essential job functions, and/or (c) assessments of KSAOs deemed necessary to perform essential job functions.

Under the ADA of 1991 (and the prior Rehabilitation Act of 1973), there is no such thing as "disability as a matter of law." The general requirements for being disabled under the law include (a) a physical or mental impairment that (b) interferes with a major life function. In addition to prongs (a) and (b), the individual must (c) be able to perform all essential job functions with

or without accommodations. These requirements are necessary to demonstrate regardless of whether the impairment is current or past or if the employer mistakenly believes an individual is disabled. As a result, disabilities fall within an interval between prongs (b) and (c) in which individuals with minor or temporary impairments cannot demonstrate interference with major life functions, and individuals with extremely severe impairments may not be able to perform all essential job functions, even with accommodations.

The first step in determining if a disabled person requires accommodations is determining whether there is a nexus between a physical or mental impairment and essential job functions. A person with one leg is clearly disabled under the ADA, but it is unlikely that accommodations beyond access to the workplace are required if the job is computer programmer. If there is a nexus, the employer should next meet with the applicant or employee and together explore potential accommodations to overcome the barrier implied by the disability. If no such accommodations are possible, the individual, unfortunately, faces an insurmountable yet legal barrier to employment.

If assessment is involved in the selection process, it is important to accommodate applicants with special needs. For example, if a paper-and-pencil or computer-presented format is used and the construct of interest is "good judgment," it may be important to allow applicants with limited vision to have the test questions read to them.[17] More generally, it is good practice to incorporate KSAOs needed to perform essential job functions in the testing process. For example, if a job or critical tasks can be performed without undue concern for the passage of time, it may be inappropriate to use demanding time limits for a test to be taken by an individual who claims a learning disability related to speed of reading or processing, as this could alter the underlying measurements or distributions of the test.[18]

One other point is worth noting: Until 2014, soliciting disability status from applicants before a job offer was prohibited by the ADA. In 2014, Section 503 of Rehab-73 was updated in various ways, including the requirement that federal contractors meeting 50-employee/$50,000 contract thresholds are required to solicit disability status information from applicants pre-offer and post-offer.

As noted by Pryor, Dunleavy, and Cohen (2014), there was immediate concern with the new regulations around whether the pre-offer solicitation of disability status violated the ADA. Similar regulations were proposed in 1996, and at that time the EEOC provided a letter stating that pre-offer solicitation would be a violation under general ADA provisions. However, in 2014, the EEOC provided a letter stating that when federal contractors are required to solicit this information to comply with a federal regulation, it will not be violating the ADA/ADAAA. It remains to be seen whether this will provide a legal safe haven for contractors if they are challenged. Regardless, the availability of applicant disability self-identification may provide selection researchers with fresh opportunities to conduct research that was impossible before these new regulations.

There are other interesting aspects of the updated regulations. For example, employee disability status must be solicited every five years. Contractors are still required to "periodically" review mental and physical job qualifications and personnel processes, although not annually. Additional new requirements include written documentation of outreach and recruitment efforts as well as a utilization analysis with a goal of 7% employment of individuals with disabilities. Contractors must assess whether they have a gap between the 7% goal and actual employment in each affirmative action job group. If there is a gap, they must strive to eliminate the gap with focused outreach and recruitment efforts, not a quota. As such, not meeting the goal does not necessarily mean violating the regulations. At the time this chapter was written, little enforcement data were available related to the new regulations.

Exemplar Case Law Related to Principle 13

The individual approach to defining disability was affirmed in three 1999 Supreme Court rulings: *Sutton v. United Air Lines* (1999), *Murphy v. UPS* (1999), and *Albertsons v. Kirkingburg* (1999). For example, in *Kirkingburg*, the Supreme Court ruled:

> This is not to suggest that monocular individuals have an onerous burden in trying to show that they are disabled. . .. We simply hold that the Act requires monocular individuals . . . to prove a disability by offering evidence that the extent of the limitation in terms of their own experience, as in loss of depth perception and visual field, is substantial.

In other words, *Kirkingburg* could have proven he was disabled within the meaning of the law, but he did not, assuming amblyopia is a disability as a matter of law.

In the other two cases, the Supreme Court ruled that impairments must be evaluated with mitigation (eyeglasses for visual impairments in *Sutton* and high blood pressure medication for hypertension in *Murphy*). In an extension of the ruling in *Kirkingburg*, the *Murphy* Court ruled:

> Murphy could have claimed he was substantially limited in spite of the medication. Instead, like Kirkingburg, [he] falsely assumed that his impairment was a disability as a matter of law.

Taking advantage of this "advice," plaintiffs subsequently proved disability despite medication, as for example, in *EEOC v. JH Routh Packing* (2001) (seizures only partially controlled with epilepsy medication) and *Lawson v. CSX* (2001) (debilitating side effects of insulin medication for diabetics). It should be noted that in the ADA Amendments Act of 2008, Congress changed the rules for mitigating measures (except for eyeglasses), thus reversing the *Albertsons v. Kirkingburg* (1999) rulings. This ruling does not, however, alter the principle of assessing impairment on a case-by-case basis.

Examples of insurmountable barriers include *Southeastern Community College v. Davis* (1979) (a deaf woman excluded from nursing school), *Treadwell v. Alexander* (1983) (a heart patient who cannot perform all-day foot patrols excluded from park ranger job), and *Miller v. Illinois* (1996) (a blind person who could perform some but not all essential functions of corrections officer). However, it is critical that the job functions in question are essential, as, for example, in *Stone v. City of Mt. Vernon* (1997) (a paraplegic former firefighter was refused a desk job because he could not fight fires in emergencies).

When accommodations are possible, plaintiffs have a duty to inform potential employers, and both parties have a duty to "flexibly interact" to seek accommodations. Examples of failure to notify include *Hedberg v. Indiana Bell* (1995) (notification of fatigue syndrome after termination) and *Taylor v. Principle Financial* (1996) (notification of bipolar disorder after a poor performance evaluation). Examples of employee failures to flexibly interact include *Beck v. University of Wisconsin Bd. of Regents* (1996) (the employee refused a request for medical records to identify accommodations) and *Grenier v. Cyanamid Plastics* (1995) (the employee refused a request for psychiatric information). Examples of employer failures to flexibly interact include *Bultemeyer v. Fort Wayne* (1996) (the employer ignored a request by a psychiatrist for reassignment), *Feliberty v. Kemper Corp.* (1996) (the employer falsely assumed that a medical doctor can design his own accommodations), *Whiteback v. Vital Signs* (1997) (the employer ignored a request for a motorized cart because "it wouldn't look right"), and *Dalton v. Suburu-Izuzu* (1998) (the employer ignored a request for step stools and guard rails without discussion).

Mistakes in assessment include *Stutts v. Freeman* (1983), in which a dyslexic applicant failed the General Aptitude Test Battery for a job (heavy truck operation) that did not require reading skills. On the other hand, in *Fink v. New York City* (1995), the city was not liable when accommodations for blind applicants (readers and interpreters) did not result in passing scores on a civil service exam.

CONCLUSIONS

The term "law" in the selection context generally has two separate but related meanings. There is statutory law embodied in the Civil Rights Acts, ADA, ADEA, and similar federal statutes that we have discussed above. There is also "case law," which is embodied in the opinions of various levels of the federal judiciary (trial, appeals, and Supreme Courts). The latter interprets the former from the legal perspective. Selection practitioners must be aware of both aspects of

"the law." They must be aware of the statutory requirements as well as how judges have interpreted these requirements. Statutory statements of the law seldom recognize the specific contributions of I-O psychology to selection practices. Judges and other EEO stakeholders, on the other hand, often cite the testimony of I-O psychologists and the standards by which selection practice is evaluated (e.g., UGESP, SIOP *Principles*). In this chapter, we have attempted to bring together practice, statutory law, and case law as a way of educating practitioners. Other chapters provide more detailed descriptions of practices. We provide a legal context for many of those practices. Context matters, and toward that end we remind I-O practitioners that (a) professional judgement cannot be removed from the equation and (b) talking to the appropriate legal counsel may be a useful approach to getting ahead of these issues.

NOTES

1. Our colleague, co-author, and friend Jim Outtz passed away shortly before this chapter was completed. The field of I-O psychology lost a leader, scholar, and model scientist-practitioner that day, and we are grateful for having had the opportunity to work with and learn from Jim. We will miss our friend, as will the rest of the field.
2. We note that the legal context is often very complex, as are evidentiary standards related to technical matters. We are not intending to give legal advice in this chapter and recommend that readers consult legal counsel if they are dealing with any of the issues discussed in this chapter.
3. Interested readers are referred to Sady, Aamodt, and Cohen (2015) for a review of recent pay equity issues and enforcement.
4. Obligation to Solicit Race and Gender Data for Agency Enforcement Purposes, 70 Fed. Reg. 58946 (Oct. 7, 2006) (codified at 41 C.F.R. pt. 60–1). This final rule went into effect on February 6, 2006.
5. 41 C.F.R. § 60–1.3.
6. For a more detailed view of controversies in adverse impact measurement, please refer to Dunleavy, Morris, and Howard (2015).
7. Again, for more detail on these statistical tests, please refer to Dunleavy, Morris, and Howard (2015).
8. Practical significance is a well-established notion in the social scientific community. For example, in the most recent *Publication Manual of the American Psychological Association* (2010), a failure to report effect sizes (as practical significance measures) is considered a defect in the reporting of research: "No approach to probability value directly reflects the magnitude of an effect or the strength of a relation. For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relation in your Results section."
9. As discussed by Murphy and Jacobs (2012), the use of the term "standard deviation" is somewhat misleading, because this phrase refers to a descriptive measure of the variability of a distribution. The more appropriate term here is "standard error," which describes the variability in a test statistic due to random fluctuation across samples.
10. An OFCCP Statistical Standards Group (1979) endorsed the absolute difference as a practical significance measure.
11. We note the complexity of the topic. From an I-O psychology perspective, consideration of alternatives could relate to different measurement methods of the same construct, measuring alternative constructs, measuring additional constructs, or implementation characteristics such as weighting and score use (e.g., cut score, banding, rank order, data point, compensatory versus multiple hurdle approaches, etc.). What is or is not reasonable from a legal perspective is a different question. We also note the complexity of defining what is equally valid, which is a complicated enough notion within the same validation strategy (e.g., comparing two criterion studies), never mind across different validation strategies (e.g., a content and a criterion study).
12. Note that transportability is a concept that originated in the *Uniform Guidelines*, but was framed as a substrategy within the criterion validation framework and limited to the transport of only criterion research.
13. In the BFOQ defense, it must be proven that it is reasonably necessary to exclude all members of a class for the business to succeed. In effect, proof of adverse impact based on height and weight requires a similar proof—that it is necessary to exclude all or most people based on such criteria.
14. Those readers interested in a comprehensive discussion of those facts are directed to the DCI Consulting website (www.dciconsult.com), where Gutman has written the following three blogs:

    (1) http://dciconsult.com/supreme-court-to-review-fisher-v-university-of-texas-another-test-of-grutter-v-bollinger-2003/

(2) http://dciconsult.com/supreme-court-punts-in-long-awaited-ruling-in-fisher-v-university-of-texas/
(3) http://dciconsult.com/5th-circuit-declines-en-banc-review-of-fisher-v-university-of-texas/

15. We note that the issue of "self-critical analyses" is complex and that in some instances such results could be used against an employer. We suggest that readers consult their legal counsel and read McPhail (2005).
16. The U.S. Supreme Court has never defined what constitutes a "substantial difference," but lower courts have used statistical significance of Chi Square and/or Fisher Exact Tests.
17. We note that whether this change is (1) an accommodation where the construct being measured is still the construct of interest, or (2) a modification changing the construct being measured is often a complex question to answer.
18. We note the complexity of such decisions, particularly when considering whether scores from an accommodated administration can be reasonably compared with scores from an unaccommodated administration. This example is further complicated by the potential legal consequences of flagging accommodated scores.

REFERENCES

American Psychological Association (2009). *Publication manual of the American Psychological Association (6th Ed.)*. Washington, DC: American Psychological Association.

American Statistical Association. (2016). P-values under question. *Psychological Science Agenda*. Retrieved March, 2016 from http://www.apa.org/

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561–590.

Cohen, D. B., Aamodt, M. G., & Dunleavy, E. M. (2010). *Technical advisory committee report on best practices in adverse impact analyses*. Washington, DC: Center for Corporate Equality.

Cohen, D., & Dunleavy, E. M. (2009). *A review of OFCCP enforcement statistics: A call for transparency in OFCCP reporting*. Washington: Center for Corporate Equality.

Cohen, D., & Dunleavy, E. M. (2010). *A review of OFCCP enforcement statistics for fiscal year 2008*. Published independently by The Center for Corporate Equality at www. cceq. org.

De Corte, W. (1999). Weighing job performance predictors to both maximize the quality of the selected work-force and control the level of adverse impact. *Journal of Applied Psychology, 84*, 695–702.

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.

Dunleavy, E. M., & Gutman, A. (2011). An update on the statistical versus practical significance debate: A review of Stagi v Amtrak (2010). *The Industrial-Organizational Psychologist, 48*, 121–129.

Dunleavy, E. M., Morris, S., & Howard, L. (2015). Measuring adverse impact in selection decisions. In K. Sady & C. Hanvey (Eds.), *HR Practitioner's Guide to Legal Issues in Organizations: Research Methods for Practical Problems* (pp. 1–27). New York, NY: Springer.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Justice, & Department of Labor. (1978). *Uniform guidelines on employee selection procedures. 29 CFR, 1607*.

Esson, P. L., & Hauenstein, N. M. (2006). Exploring the use of the four-fifths rule and significance tests in adverse impact court case rulings. In *21st annual conference of the Society for Industrial and Organizational Psychology*, Dallas, TX.

Gutman, A., Koppes, L. L., & Vodanovich, S. J. (2010). *EEO law and personnel practices (3rd ed.)*. New York: Routledge.

Hattrup, K., Rock, J., & Scalia, C. (1997). The effects of varying conceptualizations of job performance on adverse impact, minority hiring, and predicted performance. *Journal of Applied Psychology, 82*, 656–664.

Jacobs, R., Murphy, K., & Silva, J. (2013). Unintended consequences of EEO enforcement policies: Being big is worse than being bad. *Journal of Business and Psychology, 28*(4), 467–471.

Landy, F. J. (Ed.) (2005). Phases of employment litigation. In *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 3–19). San Francisco, CA: Jossey Bass.

McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). Encouraging debate on the uniform guidelines and the disparate impact theory of discrimination. *Industrial and Organizational Psychology, 4*(4), 566–570.

McPhail, S. M. (2005). Auditing selection processes: Application of a risk assessment model. *The Psychologist-Manager Journal*, *8*(2), 205–221.

Morris, S., Dunleavy, E. M., & Howard, E. (2015). Measuring adverse impact in employee selection decisions. In C. Hanvey & K. Sady (Eds.), *Practitioner's guide to legal issues in organizations* (pp. 1–27). New York, NY: Springer.

Murphy, K. R., & Jacobs, R. R. (2012). Using effect size measures to reform the determination of adverse impact in equal employment litigation. *Psychology, Public Policy, and Law, 18*(3), 477–499.

Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests. *Personnel Psychology, 50*, 823–854.

Outtz, J. L. (2007). Less adverse alternatives: Making progress and avoiding red herrings. *The Industrial-Organizational Psychologist, 45*(2), 23–27.

Outtz, J. L. (Ed.) (2010). *Adverse impact: Implications for organizational staffing and high stakes selection*. New York: Routledge.

Pryor, K., Dunleavy, E. M., & Cohen, D. (2014). Funny you should mention it: New disability EEO/AA regulations finalized for federal contractors. *Industrial and Organizational Psychology, 7*, 220–224. doi: 10.1111/iops.12135

Reynolds, D. H. (2004). EEOC and OFCCP guidance on defining a job applicant in the Internet age: SIOP's response. *The Industrial-Organizational Psychologist, 42*(2), 127–138.

Roth, P. L., Bobko, P., & Switzer, F. S., III. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology, 91*(3), 507–522.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707–722.

Sady, K., Aamodt, M. G., & Cohen, D. (2015). Compensation equity: Who, what, when, where, why, and how. In *Practitioner's guide to legal issues in organizations* (pp. 249–282). New York, NY: Springer.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719–730.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author. Reprinted with permission.

Wingate, P. H., Thornton, G. C., III., McIntyre, K. S., & Frame, J. H. (2003). Organizational downsizing and age discrimination litigation: The influence of personnel practices and statistical evidence on litigation outcomes. *Law and Human Behavior, 27*, 87–108.

Zink, D. L., & Gutman, A. (2005). Statistical trends in private sector employment discrimination suits. In F. J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 101–131). San Francisco, CA: Jossey Bass.

Cases Cited

Adams v. City of Chicago, 469 F.3d 609 (7th Cir. 2006).
Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975).
Albertsons v. Kirkingburg, 527 U.S. 555 (1999).
Apsley v. The Boeing Co., No. 05–1368 (D. Kan. Dec. 9, 2013).
Baty v. Willamette Industries, Inc., 172 F.3d 1232 (10th Cir. 1999).
Beck v. University of Wisconsin Bd. of Regents, 75 F.3d 1130 (7th Cir. 1996).
Bernard v. Gulf Oil Corp., 890 F.2d 735 (5th Cir. 1989).
Berndt v. Kaiser Aluminum & Chemical Sales, Inc., 789 F.2d 253 (3d Cir. 1986).
Bew v. City of Chicago, 252 F.3d 891 (7th Cir. 2001).
Biondo v. City of Chicago, Ill., 382 F.3d 680 (7th Cir. 2004).
Borkowski v. Valley Cent. School Dist., 63 F.3d 131 (2d Cir. 1995).
Boyd v. Ozark Air Lines, Inc., 568 F.2d 50 (8th Cir. 1977).
Bradley v. City of Lynn, 443 F. Supp. 2d. 145 (D. Mass. 2006).
Bradley v. Pizzaco of Nebraska, Inc., 7 F.3d 795 (8th Cir. 1993).
Bridgeport Guardians, Inc. v. City of Bridgeport, 933 F.2d 1140 (2d Cir. 1991).
Briscoe v. City of New Haven, No. 3: 09-cv-1642 (CSH) (D. Conn. July 12, 2010).
Briscoe v. City of New Haven, 654 F.3d 200 (2d Cir. 2011).
Brunet v. City of Columbus, 58 F.3d 251 (6th Cir. 1995).
Bultemeyer v. Fort Wayne School, 100 F.3d 1281 (7th Cir. 1996).
Burlington Industries, Inc. v. Ellerth, 524 U.S. 742 (1998).
Burlington N. & S.F.R. Co. v. White, 548 U.S. 53 (2006).
Castaneda v. Partida, 430 U.S. 482 (1977).
Coates v. Sundor Brands, Inc., 164 F.3d 1361 (11th Cir. 1998).

Contreras v. City of Los Angeles, 656 F.2d 1267 (9th Cir. 1981).
Dalton v. Subaru-Isuzu Automotive, Inc., 141 F.3d 667 (7th Cir. 1998).
Davis v. City of Dallas, 777 F.2d 205 (5th Cir. 1985).
Dees v. Johnson Controls World Services, Inc., 168 F.3d 417 (11th Cir. 1999).
Detroit Police Officers Ass'n v. Young, 608 F.2d. 671 (6th Cir. 1979).
Dothard v. Rawlinson, 433 U.S. 321 (1977).
EEOC v. Atlas Paper Box Co., 868 F.2d 1487 (6th Cir. 1989).
EEOC v. J. H. Routh Packing Co., 246 F.3d 850 (6th Cir. 2001).
Ercegovich v. Goodyear Tire & Rubber Co., 154 F.3d 344 (6th Cir. 1998).
Faragher v. City of Boca Raton, 524 U.S. 775 (1998).
Feliberty v. Kemper Corp., 98 F.3d 274 (7th Cir. 1996).
Fisher v. University of Texas at Austin, 133 S. Ct. 2411 (2013).
Fitzpatrick v. City of Atlanta, 2 F.3d 1112 (11th Cir. 1993).
Frazier v. Garrison ISD, 980 F.2d 1514 (5th Cir. 1993).
Geller v. Markham, 635 F.2d 1027 (2d Cir. 1980).
Gentry v. Export Packaging Co., 238 F.3d 842 (7th Cir. 2001).
Gillespie v. State of Wis., 771 F.2d 1035 (7th Cir. 1985).
Gratz v. Bollinger, 539 U.S. 244 (2003).
Grenier v. Cyanamid Plastics, Inc., 70 F.3d 667 (1st Cir. 1995).
Griggs v. Duke Power Co., 401 U.S. 424 (1971).
Grutter v. Bollinger, 539 U.S. 306 (2003).
Guardians Ass'n of New York City v. Civil Serv., 630 F.2d 79 (2d Cir. 1980).
Gulino v. Bd. of Educ. of City School Dist. of N.Y., 907 F. Supp. 2d 492 (S.D.N.Y. 2012).
Gulino v. New York State Educ. Dept., 460 F.3d 361 (2d Cir. 2006).
Hare v. Potter, 549 F. Supp. 2d 688 (E.D. Pa. 2007).
Hayden v. County of Nassau, 180 F.3d 42 (2d Cir. 1999).
Hazelwood School Dist. v. United States, 433 U.S. 299 (1977).
Hazen Paper Co. v. Biggins, 507 U.S. 604 (1993).
Hedberg v. Indiana Bell Telephone Co., 47 F.3d 928 (7th Cir. 1995).
Horace v. City of Pontiac, 624 F.2d 765 (6th Cir. 1980).
Hyland v. Fukuda, 580 F.2d 977 (9th Cir. 1978).
Isabel v. City of Memphis, 404 F.3d 404 (6th Cir. 2005).
Jackson v. FedEx Corporate Services Inc., 518 F.3d 388 (6th Cir. 2008).
Johnson v. City of Memphis, No. 06–2052 Ma/P (W.D. Tenn. 2006).
Johnson v. City of Memphis, 770 F.3d 464 (6th Cir. 2014).
Lanning v. SEPTA, 181 F.3d 478 (3rd Cir. 1999).
Lawson v. CSX Transp., 245 F.3d 916 (7th Cir. 2001).
Leftwich v. Harris-Stowe State College, 702 F.2d 686 (8th Cir. 1983).
Lewis v. Chicago, 299 F.Supp 1357 (N.D. Ill. 2005).
Lomack v. City of Newark, 463 F.3d 303 (3d Cir. 2006).
Lopez v. City of Lawrence, No. 07–11693-GAO (D. Mass. 2014).
Madel v. FCI Marketing, 116 F.3d 1247, 1251 (8th Cir. 1997).
Meacham v. Knolls Atomic Power Laboratory, 461 F.3d 134 (2d Cir. 2006).
Meacham v. Knolls Atomic Power Laboratory, 627 F. Supp. 2d 72 (N.D.N.Y. 2009).
McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).
Miller v. Illinois, 107 F.3d 483 (7th Cir. 1996).
Moody v. Albemarle Paper Co., 474 F.2d 134 (4th Cir. 1973).
Moore v. City of Philadelphia, 461 F.3d 331 (3d Cir. 2006).
Moore v. Southwestern Bell Telephone Co., 593 F.2d 607 (5th Cir. 1979).
Murphy v. United Parcel Service, 527 U.S. 516 (1999).
New York City Transit Authority v. Beazer, 440 U.S. 568 (1979).
Oubre v. Entergy Operations, Inc., 522 U.S. 422 (1998).
Parents Inv. in Comm. Sch. v. Seattle School, 551 U.S. 701 (2007).
Parker v. University of Pennsylvania, 2004 U.S. Dist. LEXIS 17423 (2004).
Petit v. City of Chicago, 352 F.3d 1111 (7th Cir. 2003).
PGA Tour, Inc. v. Martin, 532 U.S. 661 (2001).
Police Officers v. City of Columbus, 916 F.2d 1092 (6th Cir. 1990).
Regents of University of California v. Bakke, 438 U.S. 265 (1978).
Ricci v. DeStefano, 554 F. Supp. 2d 142 (D. Conn. 2006).
Ricci v. DeStefano, 530 F.3d 87 (2d Cir. 2008).

Ricci v. DeStefano, 129 S. Ct. 2658 (2009).
Robinson v. Shell Oil Co., 519 U.S. 337 (1997).
Rudin v. Lincoln Land Community College, 420 F.3d 712 (7th Cir. 2005).
Shaw v. AutoZone, Inc., 180 F.3d 806 (7th Cir. 1999).
Smith v. City of Boston, No. 12–10291-WGY (D. Mass. Nov. 16, 2015).
Smith v. City of Jackson, 544 U.S. 228 (2005).
Southeastern Community College v. Davis, 442 U.S. 397 (1979).
Spurlock v. United Airlines, Inc., 475 F.2d 216 (10th Cir. 1972).
Stagi v. National RR Passenger Corp., 880 F. Supp. 2d 564 (E.D. Pa. 2012).
Starceski v. Westinghouse Elec. Corp., 54 F.3d 1089 (3d Cir. 1995).
Stone v. City of Mount Vernon, 118 F.3d 92 (2d Cir. 1997).
Stutts v. Freeman, 694 F.2d 666 (11th Cir. 1983).
Sutton v. United Air Lines, 527 U.S. 471 (1999).
Talbert v. City of Richmond, 648 F.2d 925 (4th Cir. 1981).
Taylor v. Principal Financial Group, 93 F.3d 155 (5th Cir. 1996).
Treadwell v. Alexander, 707 F.2d 473 (11th Cir. 1983).
United States v. City of Buffalo, 457 F. Supp. 612 (W.D.N.Y. 1978).
United States v. Commonwealth of Virginia, 454 F. Supp. 1077 (E.D. Va. 1978)
Waisome v. Port Auth. of New York & New Jersey, 758 F. Supp. 171 (S.D.N.Y. 1991).
Watson v. Fort Worth Bank & Trust, 487 U.S. 977 (1988).
Whiteback v. Vital Signs, 116 F.3d 588 (D.C. Cir. 1997).
Williams v. Ford Motor Co., 187 F.3d 533 (6th Cir. 1999).
Zaccagnini v. Chas. Levy Circulating Co., 338 F.3d 672 (7th Cir. 2003).
Zuniga v. Boeing Company, No. 02-CV-807-TCK-SAJ (N.D. Okla. Jan. 25, 2005).