CHAPTER 12

# The Importance of Valid Selection and Performance Appraisal

## Do Management Practices Figure in Case Law?

Stan Malos

It is beyond dispute that employment-related lawsuits have proliferated in recent years (Weisenfeld, 2003). Although some have argued that even high-profile class action litigation has done little to remediate past discrimination or to deter it in the future (Selmi, 2003), others continue to maintain that the quality of management practices can and should relate directly to the success or failure of discrimination claims against employers (see, for example, Schwartz & Moayed, 2001; Thrasher, 2003). It remains an open question, however, whether core personnel functions such as job analysis, validation, or performance appraisal have discernable relationships with the results of employment-related lawsuits.

When I last undertook a systematic review of the law in this area (Malos, 1998), I focused just on performance appraisals. Even then, I remarked on the daunting number of cases in which these and related management practices had become central to the outcome of employment litigation. This past experience and ongoing professional attention to the area should have prepared me for the enormity of the current round of research, but it did not; the exponential explosion in both number and magnitude of recent discrimination

cases—not to mention some highly publicized multimillion-dollar settlements involving the likes of Coca-Cola, Home Depot, State Farm, Wal-Mart, and others—proved staggering. As a consequence, it was necessary to repeatedly refine my search criteria to avoid having to assimilate a truly unfathomable amount of potentially relevant case law. Ultimately, due to sheer volume and space constraints, I chose to focus on a noncomprehensive but (I hope) instructive sample of federal appellate court decisions published within the last five years. Such decisions tend to involve legal issues of greater general applicability and precedential value than the issues dealt with in the typically fact-intensive—and often fact-specific—opinions generated by lower court proceedings.

Having said that, and mindful of the dangers in attempting to capture policy from result-driven judicial opinions (Roehling, 1993), I believe I have been able to draw some conclusions that suggest not only that the case law informs practice but also that the converse may be true as well. Because validation processes—investigating the job-relatedness of and relationships among selection and evaluation criteria—remain central to both challenging and defending employment practices, I have organized this chapter around them. I begin by addressing the role of job analysis in validating selection processes, whose results may generate adverse impact against demographic classes protected under federal law. I continue by reexamining issues with the legal defensibility of performance appraisals, whose results may generate lawsuits when they are used to deny sought-after aspects of employment such as raises, promotions, or retention during reductions in force. I conclude by highlighting the potential for an integrating framework built around the notion of procedural justice that may hold promise for harmonizing judicial and practitioner perspectives on discrimination cases. In each of these sections, I review selected research and update the implications with a discussion of recent case law.

## The Role of Job Analysis and Validation in Employment Discrimination Litigation

The validation of potentially discriminatory selection practices has figured prominently in employment litigation for at least three decades now, as has its cornerstone, job analysis. In this section I

discuss the sources of an employer's obligation to present validation evidence in defense of adverse impact claims, and the courts' treatment of such evidence in their opinions when deciding such claims. I then review recent federal cases that deal more specifically with issues involving job analysis, minimum qualifications, validity generalization, and other validation processes, with a view toward identifying aspects of those that tend to survive judicial scrutiny.

## Job Analysis and Validity: Prior Research on Their Role in Discrimination Claims

One of the earliest scholarly reviews of discrimination cases involving job analysis was done by Thompson and Thompson (1982). After noting that the need for job-relatedness (and thus by implication job analysis) was established by the U.S. Supreme Court in *Griggs* v. *Duke Power Co.* (1971) and reaffirmed in *Albemarle* v. *Moody* (1975), the authors reviewed selected federal court cases and extracted some common principles regarding legal standards for job analysis in test validation. Basing their conclusion in part on the *Albemarle* court's opinion that validation efforts in that case were deficient because "no analysis of the attributes of, or the particular skills needed in, the studied job groups" had been made, Thompson and Thompson suggested that a job analysis should be performed on the exact job for which selection is to be made; that it should include current information from individuals familiar with the job, relevant documents, and observed job performance; and that it should be conducted by an expert analyst using enough data that the sample is relevant for every job to which a particular selection device is meant to apply. Thompson and Thompson also recommended that minimum competency levels for all tasks, duties, and activities in the job be specified. They cautioned, however, that their conclusions were drawn largely from cases that involved content validation, and should thus, by implication, be applied beyond that context with due care.

Almost ten years later, Harvey (1991) reviewed some of the basic legal issues in collecting and using job analysis data. Harvey organized his review around two sets of sources: laws and court decisions and professional standards documents. In examining

laws and court decisions, Harvey noted that job analysis data have proved useful in demonstrating the job-relatedness of selection tests under federal antidiscrimination laws (for example, Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act of 1967, and the Equal Pay Act of 1963) in that such data support the inference that selection decisions have been based on predictions of successful performance rather than discriminatory motives or effects. In examining professional standards documents, Harvey noted that various editions of the American Psychological Association's *Standards for Educational and Psychological Testing* (Standards) (for example, 1985; an earlier version was used by the Supreme Court in *Albemarle*), the Society for Industrial and Organizational Psychology's *Principles for the Validation and Use of Personnel Selection Procedures* (Principles) (1987, updated in 2003), and the Equal Employment Opportunity Commission's *Uniform Guidelines on Employee Selection Procedures* (Guidelines) (1978), in Title 29 of the Code of Federal Regulations, contain similar but not identical guidance that courts have continued to rely on to varying degrees in cases involving job analysis (see, for example, *Gulino et al.* v. *Board of Education,* 2003).

The SIOP Principles appear to be relatively liberal regarding the permissibility of validity generalization, whereas the APA Standards, the EEOC Guidelines, and a number of earlier judicial opinions have taken stricter views regarding the need for job-specific analyses directed toward the exact jobs and corresponding selection devices in question (Harvey, 1991; Thompson & Thompson, 1982). The Guidelines in particular have been criticized because of their apparent rigidity regarding job analysis and validation method specificity. Nonetheless, although not legally binding, they have often been given "great deference" by the courts, as is typical for regulations promulgated by administrative agencies within their area of expertise. Moreover, failure to comply with the Guidelines may be seen to have diminished the probative value of validation evidence in the eyes of some courts (see, for example, *Association of Mexican-American Educators et al.* v. *State of California,* 2000, citing *Albemarle* and *Clady* v. *County of Los Angeles,* 1985; for a recent case in which compliance with the Guidelines was required as part of a court-ordered consent decree, see *Reynolds et al.* v. *Alabama Department of Transportation,* 2003). The Guidelines, as well as other pro-

fessional standards, are addressed more fully in Chapter Three of this book but will be summarized here as a baseline for further discussion about the evolving role of job analysis and validation in discrimination case law.

At the outset the Guidelines caution that the "use of any selection procedure which has an adverse impact on the hiring, promotion, or other employment . . . opportunities of members of any race, sex, or ethnic group will be considered discriminatory . . . unless the procedure has been validated in accordance with these guidelines" (§ 1607.3). The Guidelines' general standards provide that criterion-related, content, or construct validity studies are acceptable (§ 1607.5); whichever is used, the applicable technical standards require at minimum that "[a]ny validity study should be based upon a review of information about the job for which the selection procedure is to be used" (§ 1607.14). These standards do allow, however, that "[a]ny method of job analysis may be used if it provides the information required for the specific validation strategy used." Further flexibility is allowed in that "[u]sers may, under certain circumstances, support the use of selection procedures by validity studies conducted by other users" (§ 1607.7) where adequate evidence of job similarity and overall fairness is shown to exist.

## Job Analysis and Validity: Recent Cases Illustrating Their Role in Discrimination Claims

*Allen et al.* v. *City of Chicago* (2003) offers a recent example of a case where the court relied heavily on the Guidelines in upholding dismissal of the plaintiffs' discrimination claims on summary judgment. In that case, the latest in a lengthy series involving the promotional practices of Chicago's police department, a panel of the U.S. Court of Appeals for the Seventh Circuit considered two class action challenges by minority officers, some of whom had failed a written qualifying test and were thus not eligible for further consideration, and others of whom had passed the qualifying exam but failed either the assessment component or the *merit* component that followed.

In response to the results of earlier litigation the City of Chicago had created the Promotion and Testing Task Force, which

recommended adopting different types of tests to assess a broader range of relevant knowledge, skills, and abilities (KSAs) than had previously been the case. These tests included a qualifying test designed to evaluate the minimum levels of KSAs "needed by sergeants on their first day on the job." Those passing this test would be eligible for further consideration. The city hired outside consultants to design the promotional exam. Basing their suggestion on a job analysis, the consultants recommended a pass-fail qualifying exam, to be followed by a written assessment exercise; the results of this assessment exercise would be used to rank order successful candidates for promotion. The city then asked the consultants to add an additional merit component in accordance with task force recommendations. Noting that a "qualifying examination can test knowledge, skills and abilities, but it cannot identify police men and women who have exhibited exceptional leadership in the field," the recommendations called for a further process to "identify police officers who have demonstrated superior ability, responsibility, and dedication to police service." In response the consultants interviewed various subject matter experts and concluded that leadership, mentoring, decision making, and interpersonal relations should be considered. A nomination process was then put in place whereby command-level officers were permitted to identify up to five candidates in their chain of command whom they believed exhibited these traits. Such candidates would still have to pass the qualifying exam.

Both the written qualifying exam and assessment exercise, which involved a simulation to evaluate supervision and situational judgment skills, had been content validated in accordance with the Guidelines. In addition, command-level officers were required to attend training designed to minimize biases, stereotyping, and personal influences before nominating their subordinates for the alternative merit process. A committee of command personnel then reviewed these nominations and forwarded a select number to the police superintendent, who made a final decision regarding any such promotions. These promotions were limited to a maximum of 30 percent of total promotions and were separate from those based on the assessment exercise.

When the new qualifying test was first administered, African American candidates achieved a lower pass rate (about 73 percent)

than white candidates (about 91 percent). However, although this difference was statistically significant, it was found acceptable by the court because it did not violate the Guidelines' four-fifths rule (see the ninth circuit's 2002 opinion in *Stout et al.* v. *Potter,* in which the court disregarded an apparent violation of the four-fifths rule due to small sample size in a U.S. Postal Service promotion case). Candidates who were successful with the qualifying test were then eligible for promotion based on either the assessment exercise or the merit selection process. On the subsequent assessment exercise, however, the success rates for both African Americans and Hispanics did in fact violate the four-fifths rule when compared to the success rate of white officers, thus showing adverse impact. In contrast, the merit process resulted in no adverse impact when comparing pass rates. However, African American and Hispanic officers who did *not* achieve promotion under this merit process challenged the allegedly arbitrary constraint on such promotions that limited their number to no more than 30 percent of promotions overall. Under this constraint, African American and Hispanic officers nominated under the merit process but not promoted did turn out to be excluded at a disproportionate rate.

After reviewing the burden-shifting process set forth by the Supreme Court in *Griggs* and *Albemarle* and incorporated into the Guidelines, the seventh circuit panel found that the district court had properly dismissed both sets of claims on summary judgment. Under that process, once adverse impact has been demonstrated, the burden shifts to the employer to show that any offending components of its selection practices are job related and consistent with business necessity. The court found job-relatedness due to the validation activities undertaken by the city's consultants. The burden then shifted back to the plaintiffs to show, in accordance with the Guidelines, the existence of an equally valid but less discriminatory alternative in order to negate the business necessity of the challenged practice, which they were unable to do.

In particular, with respect to the merit process, there was no showing that allowing more than the 30 percent relative cap on such promotions would be either equally valid or less discriminatory. The same was true for the proposed alternative practice suggested by the plaintiffs of allowing merit promotions without requiring a passing score on the qualifying exam. Although the

opinion left open the possibility that such evidence might be developed with respect to future administrations of the promotional process, the court cautioned that it would consider "factors such as the cost or other burdens of proposed alternative selection devices" in determining whether they would be "equally as effective as the challenged practice in serving the employer's legitimate business goals" (quoting the U.S. Supreme Court's 1988 decision in *Watson* v. *Fort Worth Bank & Trust*; for a related case upholding the content validity of testing for promotion to lieutenant, see *Bryant* v. *City of Chicago,* 2000). Interestingly, another panel of the seventh circuit explicitly recognized the importance of racial balance in police selection procedures in *Petit et al.* v. *City of Chicago* (2003). Elsewhere along these lines, in a case also involving police promotions, the first circuit concluded in *Cotter et al.* v. *City of Boston et al.* (2003) that a violation of the four-fifths rule that *would have* occurred had the city used the results of its promotion exam in strict rank order justified preferential promotion of lower-scoring African American officers in order to remedy past discrimination.

Another instructive case in which the employer's validation efforts in accordance with the Guidelines were central to its successful defense of discrimination claims is *Williams et al.* v. *Ford Motor Co.* (1999). In that case the sixth circuit considered evidence of content validation, criterion-related validation, and validity generalization in upholding Ford's selection processes against the plaintiffs' adverse impact claims. Ford had begun using a preemployment test known as the Hourly Selection System Test Battery (HSSTB) as part of its process for screening applicants for unskilled labor positions. The test had been developed by Ford's human resource consultants to replace a state employment services test because Ford's workers were expected to rotate among assignments rather than to remain within one particular job classification and the state test was considered inadequate to assess workers' capacity to do so. The HSSTB was designed to measure abilities related to reading comprehension, arithmetic, parts assembly, visual speed and accuracy, and manual dexterity. The HR consultants conducted an extensive job analysis based on supervisory skill inventories and ratings of the relative importance of each skill set in each job category. Candidates scoring in the lowest 25th percentile on the HSSTB were excluded by Ford from further hiring consideration.

After several years Ford commissioned a criterion-related validation study of concurrent design that found a significant correlation of .30 between HSSTB scores and job performance in a sample of recently hired workers. Ford's consultants also conducted a meta-analytic validity generalization study using predictive validation data from other employers using similar tests. Ford then presented expert testimony that given the totality of the circumstances, the test appeared to have been professionally and validly developed. This conclusion was disputed by plaintiffs' expert, who felt that the underlying job analysis failed to demonstrate clear linkages between specific job requirements and actual duties performed on the job. The trial court found this latter contention unconvincing and summarily dismissed the case. The sixth circuit agreed.

The appellate court panel, after an extensive discussion of *Albemarle* and its progeny as well as the Guidelines, reviewed the burden-shifting process for adverse impact cases discussed earlier. Noting that "[n]either the case law nor the Uniform Guidelines purports to require that an employer must demonstrate validity using more than one method," the court went on to uphold the trial court's conclusion that both content- and criterion-related evidence and the meta-analytic validity generalization study (which looked at sixty-one previous studies of similar tests used to predict performance in similar jobs) were relevant to supporting the job-relatedness of the test. In particular the court commented on the adequacy of the employee sample used in the criterion study, noting that this sample was composed of employees of varying tenure, education, demography, and skill levels. Over the argument that this sample and Ford's concurrent design were unduly restricted (*compare Albemarle*) for failing to include a wider range of low scorers (as might be included, for example, in a predictive study), the court held: "The law does not require that an employer, simply in order that low scorers may be included in validation studies, hire individuals who do not pass a pre-employment test" (citing the ninth circuit's 1985 opinion in *Clady*). The court also noted that the Guidelines themselves provide that they "do not require a user to hire or promote persons for the purpose of making it possible to conduct a criterion-related study" (§ 1607.14).

It is worth observing that, although the *Williams* court did rely on compliance with the Guidelines to support the employer's

defense, it is not clear that *failure* to comply with them would have been fatal to Ford's case. In reviewing Supreme Court precedent, including both *Griggs* and *Espinoza* v. *Farah Mfg.* (1973), the sixth circuit observed: "the Court has not ruled that every deviation from any of the Guidelines automatically results in a violation of Title VII." It went on to note that the Guidelines may well be more appropriately regarded as expert opinion than as binding legal precedent and, as such, should probably be applied with "the same combination of deference and wariness" that applies to the use of expert testimony overall.

In the final analysis, then, Ford's victory in *Williams* may be based on the court's belief that the employer did just about everything it could have done to validate its preemployment test, in a genuine effort to improve the performance of its workforce. Thus, notwithstanding the plaintiffs' examples of false-negative effects (anecdotal evidence about applicants who had scored poorly on the test but had performed well elsewhere for other companies), the court found insufficient reason to overturn the lower court's dismissal of the case.

Another case that involved the role of the Guidelines in discrimination issues surrounding the validation process is *Association of Mexican-American Educators et al.* v. *State of California* (2000). As mentioned earlier, the ninth circuit majority in that case reaffirmed the deference to which they felt the Guidelines were entitled. The case involves yet another class action brought by unsuccessful minority candidates, this time those excluded from pursuing the education profession because they had failed to pass the California Basic Education Skills Test (CBEST).

The CBEST, like some of the other qualifying exams considered here, is a pass-fail written test. It assesses reading, writing, and mathematical abilities with multiple-choice subtests. Candidates can pass the test by achieving minimum acceptable scores on each of the three subtests or a minimum overall score derived from compensatory averaging of the subtests. Historically, a disproportionate number of minority candidates had failed the test. In this action, such candidates attacked the use of the test based on the state's failure to adopt a less discriminatory process. Like the plaintiffs in *Allen,* they were unsuccessful with this argument at trial, where the district court found that the test was a valid measure of job-related skills, that the minimum passing score reflected rea-

sonable professional judgment about the basic skill levels required for the jobs in question, and that the evidence adduced at trial was insufficient to demonstrate the existence of some other equally effective screening device. On appeal the plaintiffs challenged these findings, in part because of the court's reliance on advice from an expert who did not prepare a written report and was not subject to cross-examination at trial.

After concluding that Title VII, and thus the Guidelines' validation requirements, applied to the CBEST, the ninth circuit reaffirmed a three-step process for determining the job-relatedness of a selection device, a process previously set forth by the court in *Craig* v. *County of Los Angeles* (1980). That process requires that the employer (1) specify the particular trait or characteristic being measured, (2) determine that this trait or characteristic is an important element of work behavior, and (3) demonstrate by professionally acceptable methods that the selection device is predictive of or significantly correlated with the element of work behavior identified. Basing its conclusion on the expert's use of the "pooled judgments" of knowledgeable persons such as job incumbents about the relevance of the skills tested by the CBEST, the court found no error in determining that these criteria were satisfied based on the professional standards of the time. Although acknowledging that the district court had been somewhat critical of certain aspects of the validation process (for example, "matching" test questions to skills purportedly assessed on the CBEST based on the content of textbooks used in California schools), the ninth circuit took care to mention that validation studies "are by their nature difficult, expensive, time consuming and rarely, if ever, free from error." Thus the court found "evidence—even if not overwhelming evidence—that the development and evaluation of the CBEST were appropriate and that the test measures the types of skills that it was designed to measure." This comment would seem to support the notion that judges will be sympathetic to employers who appear to attempt in good faith to comply with validation requirements, even if the evidence they adduce in this regard is not as strong as it might be.

Finally, the court found no need for the CBEST's minimum passing score to be separately validated beyond the efforts already made with regard to the test overall. Rather the court relied on the Guidelines' provision that "where cutoff scores are used, they should normally be set so as to be reasonable and consistent with

normal expectations of acceptable proficiency within the work force" (§ 1607.5). Thus, under the rather liberal "clearly erroneous" standard applicable to appellate review of trial court factual findings, the ninth circuit found the record adequate to support the district court's decision (for a similar outcome at the trial court level that relied in part on this analysis, see *Gulino et al.* v. *Board of Education,* 2002, 2003, in which the court ultimately upheld New York State teacher certification requirements as sufficiently job related despite the lack of "formal" evidence of their validity).

Like other appellate cases that have considered the propriety of minimum qualifications and cut scores, *Association of Mexican-American Educators et al.* generated multiple opinions dissenting at least in part from the majority's conclusions regarding the necessary level of performance that successful passage of a test should be designed to predict. Another example can be found in *Lanning et al.* v. *Southeastern Pennsylvania Transportation Authority* (SEPTA) (1999), a third circuit case that involved an aerobic capacity test, administered prior to the start of training for transit police positions, that was found to adversely impact women candidates. On remand to the district court following the third circuit's earlier 1999 decision (Lanning I), SEPTA prevailed when it was allowed to supplement the factual record with further evidence that its cut score, a maximum time of twelve minutes on a mile-and-a-half running test, had been "shown to measure the minimum qualifications necessary for successful performance of the job in question"—the legal standard for validating a discriminatory cut score adopted by the court in Lanning I (compare the Uniform Guidelines' standard in § 1607.5H, relied on by the ninth circuit in *Association of Mexican-American Educators et al.,* that cut scores "should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force"). The original *Lanning* result, and the Lanning I legal standard underlying it, were thus upheld by the 2002 majority in Lanning II (an opinion joined only by a judge appointed to replace a member of the original three-judge panel who had died prior to the opinion's being rendered).

Although the majority opinion in Lanning II was highly deferential to the employer, both the definition of *minimum qualifications* and the validation methodologies used to justify them were hotly contested in the dissenting opinion, suggesting that the legal standard for proper validation remains at least partially unsettled. For

example, the dissent noted that the aerobic capacity test was designed to screen for better physical fitness and thus argued for a more lenient standard of review that would allow a higher cut score in testing for jobs related to public safety. Nonetheless the results of these cases overall seem to suggest that some judges may reach a point where the economies of further validation efforts do not support requiring anything more of employers who appear to have "suffered" enough in justifying their selection devices. If so, then there may well be a trend away from the more specific job analysis and related validation requirements recommended by Thompson and Thompson (1982) and based on their review of the case law at the time.

A further example of this possible trend toward court acceptance of less rigorous but good faith validation efforts by the employer can be found in *Meeker et al.* v. *Merit Systems Protection Board* (2003). The case was brought by unsuccessful candidates for administrative law judge (ALJ) positions who challenged a decision by the director of the federal Office of Personnel Management (OPM) to deviate from strict adherence to the OPM's own applicable employment regulations. In that case the U.S. Court of Appeals for the Federal Circuit ultimately allowed a government employer to dispense with regulatory requirements for a professionally developed, job-specific job analysis that at least echo—and are perhaps even more exacting than—those in the Guidelines.

In order to alleviate unduly high failure rates the OPM had made a series of revisions to its applicant scoring system that made it easier for candidates to qualify. Whereas the previous system determined candidate eligibility based on a validly weighted combination of scores on relevant legal experience, panel interviews, written materials, and reference inquiries, the revised system made anyone who met a minimum experience requirement and merely *completed* the same four-part exam threshold qualified for consideration. At this stage the impact of performance on the four exam dimensions became compressed, statutory veteran preferences gained greater relative weight than had previously been the case, and Meeker and other nonveteran applicants were excluded from further consideration. Meeker challenged this result, citing government regulations applicable to the hiring process that required (1) a job analysis identifying the necessary duties, KSAs, and candidate qualifications for the position; (2) a professionally developed "employment practice" for the recruitment, measurement, ranking, and selection of

individuals (so that a necessary relationship between the employment practice and job performance would thereby be demonstrated); and (3) nondiscrimination on the basis of race, color, religion, sex, age, national origin, partisan political affiliation, or "any other nonmerit factor" ("Employment Practices: Basic Requirements," 2004). In a class action appeal to the Merit Systems Protection Board, Meeker prevailed. On review, the federal circuit reversed.

Noting that the OPM's director is authorized to grant a variation from the strict letter of OPM regulations to avoid "practical difficulties and unnecessary hardships" so long as the variation is "within the spirit of the [OPM] regulations, and the efficiency of the Government and the integrity of the competitive service are protected" ("Civil Service Regulations," 2004), the court cited practical difficulties with revising the final ratings of more than 1,700 candidates who had already been named to the ALJ placement register. The court also found experience level to be a traditionally important threshold selection criterion, and the resulting scale compression and relative reweighting of the remaining criteria not inconsistent with the spirit of the regulations and integrity of the process. In short, despite the fact that the OPM admitted on the record that it did not have any of the job analysis data that would be required by its own regulations, it was permitted to use an unvalidated revised system in place of its prior, validated selection method. Admittedly, this case involves a specific fact situation and the fairly lenient "abuse of discretion" administrative standard of review. Nonetheless, it does seem anomalous that the "practical difficulties" and "hardships" of validating the revised system were used to justify deviation from regulatory requirements whose purposes undoubtedly include preserving the integrity of the hiring process.

## Summary of Cases Involving Job Analysis and Validity

The results of *Meeker* and the other cases discussed earlier (summarized in Table 12.1) provide further arguable evidence of evolving judicial flexibility toward job analysis and validation requirements if and when fundamental fairness can be said to have remained intact. This issue will be revisited later (in the section discussing procedural justice as a possible integrating framework). However, a final point should be made before leaving this topic.

**Table 12.1. Summary of Recent Federal Appellate Cases Involving Job Analysis and Validity.**

| Case | Issue | Outcome and Reasons |
|---|---|---|
| *Allen et al.* v. *City of Chicago* (7th Cir. 2003) | Defensibility of multipart police promotional process that adversely impacted minority candidates | Summary judgment for the employer upheld due to adequate job analysis and content validation in accordance with the Uniform Guidelines. |
| *Williams et al.* v. *Ford Motor Co.* (6th Cir. 1999) | Defensibility of multipart ability test for rotational production workers that adversely impacted minority candidates | Summary judgment for the employer upheld due to job analysis, content and criterion validation, and validity generalization in accordance with the Uniform Guidelines. |
| *Ass'n of Mexican-American Educators et al.* v. *State of California* (9th Cir. 2003) | Defensibility of basic skills test (reading, writing, and mathematics) for public school teaching eligibility that adversely impacted minority candidates | Trial judgment for the state upheld due to adequate—"if not overwhelming"—evidence of reasonable test development and cut scores in accordance with the Uniform Guidelines. |
| *Lanning et al.* v. *SEPTA* (3d Cir. 1999, 2002) | Defensibility of minimum aerobic capacity and maximum running time test for transit police officer trainees that adversely impacted women candidates | Trial judgment for the employer upheld due to adequate evidence that the test measured "minimum qualifications necessary for successful performance of the job in question." |
| *Meeker et al.* v. *Merit Systems Protection Board* (Fed. Cir. 2003) | Defensibility of unvalidated revised scoring system for administrative law judge positions that revalued veteran preferences and thus adversely impacted nonveteran candidates | Appeal in favor of plaintiffs overturned, thus upholding employer's decision to disregard its own specific regulations regarding job analysis and professional test development. |

In several of the adverse impact cases discussed here, once the employer was able to successfully rebut a four-fifths rule violation with acceptable validation evidence, plaintiffs' arguments regarding possible alternative selection methods that might be equally effective with less adverse impact were given very short shrift by the courts. A likely explanation is that such plaintiffs were seldom able to present convincing evidence of *actual* methods that were equally effective and instead presented arguments or speculation about the possibility that such methods might exist (see, for example, *Gulino,* discussed earlier, in which the New York trial court rejected a teacher certification method used in Connecticut as insufficiently supported in this regard). Of course, given judges' willingness to factor in cost-benefit analyses in comparing the "effectiveness" of proposed alternatives, in light of the Supreme Court's ruling in *Watson* v. *Fort Worth Bank & Trust,* this is a tough standard to meet in actual practice. Indeed, it would be the rare plaintiff, even in a class action, who would be able to marshal the resources and access to necessary company data sufficient to conduct his or her own validation study of a different selection system within the time parameters of a typical lawsuit. Were plaintiffs to complain to the court that this sets up a playing field that is tilted against them, they might well be met with the ironic response that the very burden of obtaining such evidence shows the alternative procedure to be *not* equally effective. In any event, if future plaintiffs hope to achieve greater success in rebutting business necessity by proposing an alternative procedure, they will likely need to improve their showing in this regard.

## The Role of Performance Appraisal in Employment Discrimination Litigation

Whereas selection has to do with predicting the performance of possible future hires, performance appraisal involves assessing both the actual performance of those hired and the efficacy of the prediction process. When either proves problematic or outdated or when economic conditions so require, appraisal results are typically relied on to modify or deny some desired benefit of the employment relationship. It is thus not surprising that appraisals generate disagreements with and challenges to their effects when

an employee views the results as unfavorable. This has been particularly true in an environment of global competition, mergers and acquisitions, reorganizations, temporary layoffs, and permanent reductions in force (RIFs), in which performance is often asserted as the reason for adverse action (if indeed a reason is given) but discriminatory purposes are thereafter alleged (see, generally, Martin, Bartol, & Kehoe, 2000; Williams, Slonaker, & Wendt, 2003). In this section, I begin again with a discussion of selected prior research on the legal defensibility of performance appraisals, followed by a discussion of recent appellate case law.

## Performance Appraisal: Prior Research on its Role in Discrimination Claims

Numerous reviews of the case law involving legal implications of performance appraisals have been written over the years. These date back at least as far as Lubben, Thompson, and Klasson (1980) and Klieman and Durham (1981) and continue forward to Martin et al. (2000), the latest update by Martin and his colleagues of several previous reviews. A fuller listing of these reviews can be found in my previous treatment of the topic (Malos, 1998, pp. 79–83). From this literature and analysis of the case law, I extracted some recommendations for improving the legal defensibility of performance appraisals. They include using specific, objective, behavioral, and job-related criteria, as well as ensuring procedural standardization, documentation, employee notification, opportunity to correct deficiencies, and other aspects of procedural justice. Along similar lines, Werner and Bolino (1997) found empirical support for the roles of job analysis, reviewability, and other aspects of due process in explaining appellate court decisions involving performance appraisals (this work will be revisited later).

Although reviews of the case law may be time-bound to the decisions on which they are based, at least two of the observations I made previously hold true today. First, the proliferation of cases involving performance appraisals seems to have continued virtually unabated; hundreds of such cases can be found at various levels in the reported federal court decisions alone in the five-year period since my last review. Second, unlike those involving validation processes, which tend to arise in the context of adverse impact

claims, cases dealing with performance appraisals still arise most often in the disparate treatment context, which makes it more difficult to infer broadly applicable principles from them.

This latter point is particularly true in cases that use the burden-shifting process set forth by the Supreme Court in *McDonnell Douglas Corp.* v. *Green* (1973). Under that process, plaintiffs who allege discriminatory employment actions typically offer evidence of previous favorable appraisals as part of their prima facie case to show that they were qualified for the job in question. Employers then typically counter with evidence that performance was poorer than asserted or had deteriorated over time, that performance standards had changed due to increased competitive pressures, or that an individual who *was* selected or promoted instead of the plaintiff exhibited better likely performance, thus supporting the assertion that the plaintiff's treatment was based on a legitimate, nondiscriminatory reason. Because of the predominantly factual nature of such disputes, these cases usually entail extensive court discussion of lengthy records that can turn on the credibility of witnesses, the adequacy of evidence to support inferences drawn on summary judgment or by the jury at trial, the appropriate standard of appellate review, or some combination of all these factors. As a consequence, one can sort optimistically through numerous cases that appear to "involve" performance appraisals, only to find that many, if not most, reduce to subjective disputes about the quality or value of the plaintiff's work.

An apt example in the context of race-based, hostile environment allegations can be found in *Hawkins* v. *Pepsico* (2000), in which the African American plaintiff claimed that her termination for poor performance was in reality due to discriminatory animus that led to improper lowering of her appraisal results. A panel of the fourth circuit, after an exhaustive—and exhausting—review of the evidence, affirmed summary dismissal of the case, finding that the plaintiff "has shown nothing more than a routine difference of opinion and personality conflict with her supervisor . . . we refuse to transmute such ordinary workplace disagreements between individuals of different races into actionable race discrimination." (Further anecdotal support for this kind of sentiment—judicial impatience with court actions that center on the "fairness" of everyday management decisions that have no clear linkage to

protected class status—can be found in Landy's interviews with judges in Chapter Fifteen of this volume.) For our present purposes, these cases, although interesting, provide little useful guidance about the overall propriety of performance appraisal *processes.* Thus, even though appraisals are theoretically considered *tests* subject to the Uniform Guidelines, it remains challenging to find opinions that explicitly address performance appraisal compliance with formal notions of job analysis, validation, and the like (see Malos, 1998, pp. 77–79, citing Barrett & Kernan, 1987).

## Performance Appraisal: Recent Cases Illustrating Its Role in Discrimination Claims

Conversely, there are cases that stand for fairly straightforward propositions that are nonetheless important. For example, it is now widely understood in most jurisdictions that a negative appraisal, without additional consequences, will not constitute an adverse employment action giving rise to a civil rights claim (see, for example, the eighth circuit's 2000 opinion in *Spears* v. *Department of Corrections,* in which the court rejected the claim that reduction from a rating of "highly successful" to "successful" was sufficient to fend off summary judgment in favor of the employer where no further use of the evaluation to the plaintiff's detriment was ever made). However, where there is evidence that negative evaluations allegedly tinged with discriminatory animus resulted in a reduced bonus or other pay reduction, summary judgment dismissing the case clearly is not appropriate (for example, *Russell* v. *Principi,* 2001).

Another general principle involves the historically common fact pattern whereby previously favorable appraisals are suddenly replaced by one or more unfavorable assessments just prior to a layoff. Although such changes are now often defended as arising from the adoption of more stringent performance standards, showing such a pattern will often be sufficient for the plaintiff to avoid summary dismissal, thus putting the employer to the risk and expense of trial. In such cases it is not even a very strong showing that is always required. For example, in an opinion reversing the previous dismissal of a retaliation claim brought by a secretary at a large employment law firm against one of its partners, Chief Judge Posner of the seventh circuit noted that

> It is common for supervisors to overrate their subordinates for purposes of building morale, avoiding conflict, and deflecting criticisms. . . . Not much weight can be given to [such] positive reviews. But not much does not equal zero. And by going out of his way to say nice things about the plaintiff [the defendant] made it possible for a reasonable trier of fact to infer that his later denigration of her performance was invented for purposes of the litigation . . .
>
> . . . Of course we do not hold that this is the correct interpretation of the events, only that the matter is sufficiently in doubt to require a trial [*Pryor* v. *Seyfarth et al.,* 2000, at 979–980].

For a similar finding see *Shackelford* v. *Deloitte & Touche,* a 1999 case in which the fifth circuit reversed summary dismissal of Shackelford's retaliation and race claims, noting the "tight temporal proximity" between protected activity and adverse employment action: "After four years of positive reviews, Shackelford received her first negative performance appraisal [shortly] after the class action suit was filed against D & T." Shackelford also had been listed as a witness in the case prior to her termination for "poor performance."

***Race and National Origin Cases***

Other principles of general applicability arise in the context of particular types of cases. For example, in what was apparently a national origin case (the court repeatedly referred to plaintiff's "ethnicity" but, in applying New Jersey state antidiscrimination law, was unclear whether this was a national origin or race case), the third circuit reviewed allegations that the multicomponent performance scores of Cardenas, a Mexican American, had been systematically rounded downward by white supervisors, whereas those of nonminority employees had been rounded upward. The court found this sufficient to reverse the lower court's summary dismissal (*Cardenas* v. *Massey et al.,* 2001; also relevant in this hostile environment retaliation case were allegations that a supervisor referred to Cardenas as *mojado* ("wetback") and as the "boy from the barrio," regularly inquiring whether Cardenas intended to pull out a switchblade to settle professional disagreements).

However, as foreshadowed in *Hawkins* v. *Pepsico,* not every employment dispute involving performance appraisals will give rise

to a cognizable civil rights violation. Thus in *Cullom* v. *Brown* (2000), a case involving alleged racial motivation in delaying the promotion of a fifty-five-year-old African American by the Veterans Administration, the seventh circuit rejected an allegation that the delay was attributable to unduly *favorable* performance appraisals because of his race, but for which Cullom would have qualified for a performance improvement plan, improved his performance sooner, and thus achieved an earlier promotion. The facts of the case suggest that Cullom had repeated and ongoing performance problems but that, in a misplaced effort to avoid his continuing complaints, the VA ordered that he be falsely overrated. In an opinion echoing Judge Posner's comments in *Pryor,* the court commented that Cullom's supervisors at the VA could hardly be seen to have "'retaliated' by giving him, an incompetent employee, undeserved *favorable* treatment and evaluations (and ultimately a promotion to GS-11) [when otherwise] he would have likely been demoted, placed on probation, and quite possibly terminated." In the final analysis the court ruled that the "undeniably poor policy" of "kicking someone upstairs (with more pay and a higher grade level) instead of kicking him down and possibly out" did not violate Title VII.

The foregoing case does illustrate, however, the dangers inherent in falsely evaluating someone favorably to avoid conflict in the short term, thereby creating a record that may make it difficult to discipline or terminate the employee when necessary later on. This fact pattern was illustrated some years earlier in *Vaughn* v. *Edel* (1990), a wrongful termination case in which Texaco took a hands-off approach to performance management with the plaintiff, a black attorney, out of fear of discrimination liability, but notified white employees in similar positions of performance deficiencies and afforded them the opportunity to improve. This well-meaning but misplaced conduct provided direct evidence of differential treatment, and thus discriminatory motive, based on race, thus supporting plaintiff's disparate treatment claims. A recent and highly similar fact pattern involving failure to offer a performance improvement plan (PIP) to an African American employee when such plans were offered to other, nonminority employees in similar jobs also led the sixth circuit to reverse summary dismissal in *Johnson* v. *Kroger* (2003).

***Age and Other RIF Cases***

Conversely, the eighth circuit in *Mayer* v. *Nextel* (2003), an age discrimination case, just as recently found no problem with an employer's failure to put sixty-year-old Mayer on a performance improvement plan prior to firing him for poor performance. Mayer, who had had a record of acceptable but not stellar performance as a sales manager, fared less favorably after Nextel adopted a more extensive, multifaceted appraisal mechanism with added levels of review. Mayer claimed that it showed discriminatory animus to have offered PIPs to younger sales managers but not to him. In finding this contention insufficient to fend off summary judgment in favor of the employer, the court accepted "Nextel's uncontroverted explanation [that] only managers who failed to meet quota were placed on PIPs. Because Mayer had met quota, he was not eligible for a PIP. Nextel can certainly choose how to run its business, including not offering at-will employees a PIP before termination, as long as it does not unlawfully discriminate in doing so. We refuse to sit as a super-personnel department who second-guesses Nextel's business decisions" (citing *Dorsey* v. *Pinnacle Automation,* 2002).

Another age case involving performance appraisals in which the employer prevailed, this time at trial, is *Sauzek and Koski* v. *Exxon Coal USA* (2000). In that case the age-protected plaintiffs alleged that Exxon had unfairly targeted them for inclusion in a RIF by evaluating their performance as extremely poor after twenty years of service with adequate appraisals. However, the seventh circuit found ample evidence on the record that employees both under and over forty years of age had experienced major fluctuations in their evaluations and that Exxon's rankings were not unusually harsh for older employees overall. The court thus concluded there was insufficient evidence to support an inference that the sudden drop in performance evaluation levels demonstrated underlying pretext in Exxon's assertedly performance-based termination.

Employers have met with mixed results in other cases alleging pretext in similar factual patterns, particularly those involving RIFs or other restructurings. For example, in *Thorn and Curran* v. *Sunderstrand Aerospace* (2000), the two age-protected plaintiffs were included in RIF layoffs for alleged low productivity. Thorn was able to adduce evidence that a previously favorable review was lowered

by his supervisor to avoid application of a seniority tiebreaker under which Thorn would have been retained; not surprisingly, the seventh circuit found this evidence adequate to overturn the district court's summary dismissal of Thorn's claim (for another age case that found an employer's "conscious, unexplained departure" from its own RIF procedures problematic in the context of a pretext allegation, see *Tyler et al.* v. *Union Oil,* 2002; see also *Yates* v. *Rexton, Inc.,* 2001). Curran, however, unsuccessfully asserted that the employer had wrongfully retained lesser-performing younger employees because they might contribute more to the company in the long run. In rejecting the logic of this contention, and thus any basis for overturning the summary judgment against Curran, Chief Judge Posner reasoned that

> [I]n making RIF decisions an employer is free to decide which employees are likeliest to contribute most to the company over the long haul [citations from various circuits omitted]. It would be a foolish RIF that retained an employee who was likely to quit anyway in a few months while riffing one likely to perform well for the company over a period of years. High turnover of skilled workers can be very harmful to a company. The worker who leaves may take with him trade secrets valuable to a competitor or the benefits of specialized training that the employer had given him, at some expense, in the hope of recouping the expense in the worker's superior productivity now to be enjoyed by another employer. Since younger employees tend to be more mobile than older ones, there is no basis for an inference that employers interested in the long-term potential of an employee prefer young to old [at 389].

Another recent seventh circuit case that upheld an employer's discretion to make RIF decisions according to its own business judgment, particularly where its selection processes are carefully thought out, well documented, and reviewed by legal counsel, is *Cerutti et al.* v. *BASF et al.* (2003). In that case BASF had instituted a new business plan whose purpose was to reduce the personnel headcount and "repopulate" the organization with individuals who demonstrated specific behavioral skills and attributes considered necessary for the company's future success. These criteria included the ability to "do more with less."

After offering voluntary early retirement to certain long-term employees, the company and its consultants systematically developed new job families and corresponding key competencies within family. These competencies were in turn assessed via problem-solving exercises, role plays, and targeted interviewing, all of which were done by telephone to avoid disclosing age, race, or other demographic characteristics of individual candidates. The results of these processes were subject to review by panel discussion among assessment personnel and were then forwarded to BASF's in-house legal department, which reviewed them for possible problems. Finding no statistically significant adverse impact on any protected group, retention decisions were finalized but later challenged as discriminatory based on age and in some cases race and national origin. These challenges were dismissed by the district court, an outcome affirmed on appeal.

Plaintiffs had asserted that the new business plan was a pretext for discrimination in that it disregarded their prior positive performance reviews. The seventh circuit found it unnecessary even to address this contention, finding that the plaintiffs had not satisfied their burden of establishing they were qualified, a critical component of a prima facie case under *McDonnell Douglas* (the court had already disregarded as inconsequential "stray remarks" the use of colloquial phrases such as "out with the old, in with the new" that plaintiffs asserted were direct evidence of discriminatory motivation). Noting that the primary and well-documented purpose of the restructuring was to identify and apply new performance criteria to help the company become more competitive in the future, the court held that plaintiffs had not shown they were meeting the employer's "legitimate workplace expectations" (citing its 2002 decision in *Peele* v. *County Mutual Insurance Co.*), thus rendering prior positive reviews irrelevant either to show prima facie qualifications *or* to argue pretext. In so holding, the court also reaffirmed its prior decision in *Scott* v. *Parkview Memorial Hospital* (1999), in which it observed that employers are not legally required "to prefer paper-heavy evaluations over contextual assessments by knowledgeable reviewers, or to exalt an assessment of past conduct over a prediction of future performance" (for an analogous result in a similar context see *Haywood* v. *Lucent Technologies,* 2003).

A note of caution is in order, however, lest the foregoing discussion give an impression that past performance evaluations are unimportant during RIFs. In a case that should warn against use of potentially outdated and marginally relevant appraisals, the first circuit found problematic the resurrection of poor appraisal results that had not previously been used against a black customer service representative but were later used to justify her RIF layoff (*Thomas* v. *Eastman Kodak,* 1999). This conduct was found to extend by several years the usual six-month statute of limitations applicable to discrimination claims. In line with the principle discussed earlier that a poor evaluation, without more, would not be actionable, the first circuit upheld the district court's rationale that to decide otherwise would "require a given plaintiff to file EEOC charges successfully for each performance evaluation, informal feedback from a supervisor, or office rumor [suggesting poor performance], so long as these events—even if not harmful in themselves—might be informed by racial animus and could someday contribute to a later, harmful result. This requirement would surely disrupt the American workplace . . . [by] forcing employees to 'run to the EEOC' each time they disagree with a performance evaluation."

***Retaliation Cases***

A further point should be made about some of the cases already discussed (for example, *Cullom* v. *Brown*; *Cardenas* v. *Massey et al.*; *Sauzek and Koski* v. *Exxon*). These and other cases illustrate an apparent trend in recent years for litigants to routinely assert retaliation claims in what would otherwise seem to be fairly straightforward discrimination lawsuits. This trend is probably based on a desire to increase the likelihood of obtaining higher damage awards (compensatory and punitive damages for Title VII discrimination and harassment claims are capped under the Civil Rights Act of 1991 at $300,000 per plaintiff per incident for the largest employers, but are not capped for analogous cases brought under 42 U.S.C. § 1981 or for state law claims alleging retaliation for complaining about prohibited conduct). Whatever the reasons, these cases constitute a growing category in which close temporal proximity between complaints about prohibited conduct to the employer, the EEOC, or a state human rights agency and subse-

quent adverse consequences may require an employer to defend the efficacy of its appraisal practices in multiple contexts.

An example can be found in the ninth circuit's majority opinion in *Winarto* v. *Toshiba American Electronics* (2001). In that case the plaintiff was a woman of Indonesian ancestry who successfully sued Toshiba for discrimination and harassment on a variety of bases including race, sex, national origin, disability, and retaliation for complaining about these issues to Toshiba's HR department. The retaliation claims centered on her inclusion in a RIF after previously favorable performance ratings were lowered for allegedly improper reasons not long after some of her complaints. The district court granted a motion to set aside the jury's verdict for Winarto, but the ninth circuit reversed and reinstated most of the jury's conclusions, finding the trial record adequate to support Winarto's claims. The record contains a troubling factual history of coworker abuses, including ridicule of Winarto's ancestry and accent, as well as outright physical assault, followed by cursory management responses that if believed, probably offer a blueprint for how *not* to respond to discrimination and harassment claims. However, key to the ninth circuit's opinion was the fact that Winarto, who had degrees in fields relevant to her PC support analyst position and more programming experience than most other members of her management information systems workgroup, had made numerous complaints about wrongful conduct that were not adequately addressed other than by reduced performance evaluations that conclusorily discussed "declining performance." Of course, in a harassment case even *actual* declining performance may be used to prove hostile environment if that decline occurs as a result of severe and pervasive conduct involving a worker's protected status. In any event, the majority found ample evidence that Winarto had demonstrated a prima facie case of retaliation (causal linkage between protected activities and subsequent adverse employment actions) and that Toshiba's asserted performance-based reasons for including her in the RIF were pretextual (a partial dissenting opinion takes issue with whether Winarto was held to a high enough burden of proof in the lower court's jury instructions).

To be sure, the fact patterns in retaliation cases differ widely (see, for example, *Liu* v. *Amway Corp.,* 2003, in which it was alleged that the plaintiff's appraisal ratings were lowered in retaliation for

failing to perform her duties while on extended leave under the terms of the Family and Medical Leave Act). The important point to keep in mind, however, is that an employer may successfully defend an underlying discrimination claim if its RIF criteria or other evaluation processes are upheld, only to face secondary liability for retaliation if complaints about the alleged discrimination appear causally linked with subsequent unfavorable actions (see, for example, *Fine* v. *Ryan International Airlines,* 2002; *O'Neal* v. *Ferguson Construction,* 2001; *Shannon* v. *Bell South Telecommunications,* 2002). Employers would thus be wise to scrutinize with particular care performance or other criteria used to terminate employees who have previously complained about improper conduct, whatever merit those complaints may be found to have held. The same is true for employees chosen for termination by supervisors who have been the subject of discrimination or harassment complaints in the past and whose decisions may thus more credibly be challenged based on similar allegations by other plaintiffs (see, generally, Malos, 1998, pp. 92–93, citing Cathcart, 1996).

## Summary of Cases Involving Performance Appraisal

It may appear from the foregoing discussion that employers fared less well in recent cases involving performance appraisal than they did in those involving job analysis or validity (compare Table 12.1 with Tables 12.2 and 12.3, which summarize performance appraisal cases decided for and against the employer, respectively). In numerical terms this may well be so. However, most of the appraisal cases decided for employers generated substantial closure for the defense (for example, by upholding summary dismissal of plaintiffs' lawsuits), whereas those decided against employers often only reinstated plaintiffs' claims, leaving the outcome still to be determined at trial. Put another way, although the courts have allowed some plaintiffs to "keep the ball in the air" in factual disputes that arguably require a trial to resolve, they also seem to have had no trouble affirming summary dismissal of cases in which they are asked to second-guess the business judgment of employers but can find no convincing evidence of discrimination or other wrongful conduct that compels them to do so.

**Table 12.2. Recent Federal Appellate Cases Involving Performance Appraisals: Results Favorable for the Employer.**

| Case | Issue | Outcome and Reasons |
|---|---|---|
| *Hawkins* v. *Pepsico* (4th Cir. 2000) | Defensibility of termination for poor performance where it is alleged that appraisal results were lowered due to racial discrimination motives | Summary judgment for the employer upheld due to lack of evidence that the case involved anything other than an "ordinary workplace disagreement" with a supervisor whose race differed from plaintiff's. |
| *Spears* v. *Department of Corrections* (8th Cir. 2000) | Defensibility of alleged race-based retaliation in reducing plaintiff's appraisal from "highly successful" to merely "successful" so as to have justified quitting in a constructive discharge case | Summary judgment for the employer upheld due to the well-accepted general rule that arguably negative performance appraisals, without tangible harm as a result, are not actionable. |
| *Cullom* v. *Brown* (7th Cir. 2000) | Defensibility of delay in black employee's promotion due to appraisal results allegedly too favorable, but for which employee would have known sooner that his performance had to improve | Summary judgment for the plaintiff overturned as improper due to lack of evidence of harm; retaliation not cognizable where arguably incompetent employee was promoted due to undeserved favorable rating, delayed or otherwise. |

| | | |
|---|---|---|
| *Mayer* v. *Nextel* (8th Cir. 2003) | Defensibility of failure to put age-protected employee on a PIP prior to firing him for poor performance when younger employees were placed on such plans | Summary judgment for the employer upheld due to "uncontroverted explanation [that] only managers who failed to meet quota were placed on PIPs" when the plaintiff had in fact met quota. |
| *Sauzek and Koski* v. *Exxon Coal USA* (7th Cir. 2000) | Defensibility of appraisal scores markedly lower than previous ratings where pattern just prior to a RIF was alleged to suggest possible age-related pretext | Trial verdict for employer upheld due to evidence that appraisal scores for workers both over and under 40 years of age fluctuated to similar extents, thus belying pretext in the jury's view. |
| *Thorn and Curran* v. *Sunderstrand Aerospace* (7th Cir. 2000)* | Defensibility of RIF-based layoffs of age-protected plaintiffs for alleged low productivity | Summary judgment for the employer upheld—against Curran—based on apparent judicial notice that younger workers are more turnover-prone, thus belying claim that younger workers were retained due to greater long-term value. |
| *Cerutti et al.* v. *BASF et al.* (7th Cir. 2003) | Defensibility of RIF criteria developed pursuant to new business plan designed to "repopulate" the company with those who could "do more with less" but which disregarded prior favorable performance reviews, thus allegedly showing pretext | Summary judgment for the employer upheld due to well-developed, job-related layoff criteria and plaintiffs' inability to demonstrate threshold qualifications under the new criteria, thus rendering moot any pretext allegations for lack of a prima facie case. |

*Mixed results.

**Table 12.3. Recent Federal Appellate Cases Involving Performance Appraisals: Results Unfavorable for the Employer.**

| Case | Issue | Outcome and Reasons |
|---|---|---|
| *Russell* v. *Principi* (D.C. Cir. 2001) | Defensibility of alleged discrimination in elements of compensation or promotional processes based on lower appraisal ratings | Summary judgment for the employer overturned as improper where evidence created triable issue as to whether a bonus was unfairly low due to racial favoritism. |
| *Pryor* v. *Seyfarth et al.* (7th Cir. 2000) | Defensibility of appraisal scores markedly lower than previous ratings where pattern just prior to a layoff may suggest discriminatory pretext | Summary judgment for the employer overturned as improper where evidence created triable issue as to whether such a pattern supported plaintiff's gender-based retaliation claim. |
| *Shackelford* v. *Deloitte & Touche* (5th Cir. 1999) | Defensibility of appraisal scores markedly lower than previous ratings where most recent ratings were used to justify termination for allegedly poor performance | Summary judgment for the employer overturned as improper where "tight temporal proximity" between plaintiff's complaints and firing created triable issue as to whether pattern was retaliatory. |
| *Cardenas* v. *Massey et al.* (3d Cir. 2001) | Defensibility of rank and pay differentials due to alleged adjustment of appraisal scores in opposite directions based on ratees' race or national origin | Summary judgment for the employer overturned as improper where evidence created triable issue regarding the plaintiff's claims amid allegations of epithets such as *mojado* ("wetback"). |

| | | |
|---|---|---|
| *Thorn and Curran* v. *Sunderstrand Aerospace* (7th Cir. 2000)* | Defensibility of RIF-based layoffs of age-protected plaintiffs for alleged low productivity | Summary judgment for the employer overturned as improper—against Thorn—where evidence created triable issue as to whether good review was reduced to avoid seniority tiebreaker under which Thorn would have been retained. |
| *Thomas* v. *Eastman Kodak* (1st Cir. 1999) | Defensibility of relying on unused but outdated prior poor appraisals to justify RIF-based layoff of minority employee | Lower court's rejection of employer's statute of limitations argument upheld because any complaint over negative appraisal results would not have been ripe until tangible harm occurred. |
| *Winarto* v. *Toshiba American Electronics* (9th Cir. 2001) | Defensibility of RIF-based layoff due to asserted poor performance when ratings were allegedly lowered in retaliation for complaining about discrimination | Lower court's order setting aside jury verdict for the plaintiff overturned as improper due to adequate evidence in the trial record to support plaintiff's claims. |

*Mixed results.

## Procedural Justice and Discrimination Litigation: Toward an Integrating Framework

Both procedural justice (a psychological construct) and procedural due process (a legal construct) have been discussed as possible frameworks for analyzing and improving the perceived fairness and corresponding defensibility of a variety of personnel practices (for the comparisons and contrasts among elements of these two constructs, as well as linkages among them, see Posthuma, 2003). These related concepts of procedural fairness are considered implicitly, if not explicitly, in many of the judicial opinions discussed earlier in this chapter.

They have also been addressed in various reviews and research articles, fewer of which deal with selection processes and more of which deal with performance appraisals or discrimination claims in general (for a broader recent treatment of this topic, see Landy & Conte, 2004).

For example, in a study of applicant reactions to selection practices, Bauer, Truxillo, Sanchez, Craig, Ferrara, and Campion (2001) investigated and supported the validity of an instrument designed to capture aspects of procedural justice developed in previous research (Bauer and her colleagues found that job-relatedness, among other things, would likely be related to perceptions of the overall fairness of an employer's selection procedures). And in a study of court cases involving employment interviews, Williamson, Campion, Malos, Roehling, and Campion (1997) found that elements of interview structure such as job-relatedness, standardization, and other subdimensions of procedural fairness figured significantly in judicial opinions sustaining the selection practices of employers against both disparate treatment and adverse impact claims, although to differing extents.

Similarly in the performance appraisal context, Werner and Bolino (1997) found significant effects for elements of procedural fairness, due process, and accuracy in federal appellate cases decided for the employer, even though formal mention of validation processes seldom arose. In addition to this empirical study, a number of sources already mentioned (for example, Landy & Conte, 2004; Malos, 1998; Martin et al., 2000; Williams et al., 2003) have also cited procedural justice as likely to increase perceptions

of accuracy and fairness, thus decreasing the likelihood of discrimination litigation involving an employer's personnel practices.

Other more integrative models have been proposed that link various justice-based concepts with desirable organizational outcomes: for example, increased perceived fairness, increased employee desire to improve performance, and decreased incidence of discrimination claims (Flint, 1999; Goldman, 2001). These models seem conceptually sound as far as they go, and offer promising guidance for improving personnel practice effectiveness and legal defensibility, but as yet have received only preliminary empirical support. In addition these models do little to balance the already heavy focus on appraisal procedures, which are probably more amenable to direct employee scrutiny or input than are job analysis or validation processes. Further development and extension of such models will thus be needed before they are likely to prove useful in driving advances in professional standards and practices for job analysis, validation, and performance appraisal, as well as in encouraging judicial acceptance of these standards and practices in court.

## Conclusion

It remains true according to the Guidelines and other professional standards, and according to the appellate decisions involving them, that "under no circumstances will the general reputation of a test . . . its author . . . or [unsubstantiated] reports of its validity be accepted in lieu of evidence of validity" as an adequate defense of challenged personnel practices (*Williams et al.* v. *Ford Motor Co.,* at 540–541). Indeed, the case law involving job analysis and validity, and to a certain extent that dealing with performance appraisal, does appear to support a trend toward explicit judicial consideration of the work of industrial-organizational psychologists and human resource professionals in determining the appropriate legal standards in this regard.

For example, with respect to job analysis and validity, the cases summarized in Table 12.1 reflect substantial appellate court reliance on expert testimony for evidence regarding accepted methods of developing and validating various selection devices in accordance with applicable professional standards. That said, in

resolving the often epic battles of the experts that have led to findings for the defendant of late, even in the absence of technical or formal validation, some judges appear to have reached a point where they simply cannot bring themselves to burden the employer any further in trying to ensure that tests adequately screen prospective hires based on predicted performance rather than impermissible demographic factors. To the extent that this may reflect something of a human, if not legal, presumption in favor of employers who have undertaken reasonably professional validation efforts, plaintiffs in these cases should realize that they increasingly have their work cut out for them in practical terms.

With respect to performance appraisal, professionally developed and job-related criteria, as well as various aspects of due process and procedural justice (for example, notice of performance deficiencies, opportunities to correct those deficiencies, and reviewability of adverse decisions), have also figured prominently in at least some of these cases (see, for example, *Cerutti et al.* v. *BASF et al.* and other cases summarized in Table 12.2). If these examples represent a trend, then it is probable that litigants who pay more attention to such matters will be better able to identify and adopt effective strategies for persuading the courts to find in their favor.

In any event the interplay of professional and judicial consideration of these issues will most likely continue, and we can hope that it will lead to more widespread adoption of valid, job-related selection procedures and performance appraisal processes. Such developments would serve to enhance both organizational effectiveness and employee development while also providing more reliable guidance in the context of litigation when the inevitable lawsuits do in fact occur. To this end the ongoing evolution of legal and practical analysis will be most welcome.

## References

### *Books and Articles*

Barrett, G. V., & Kernan, M. C. (1987). Performance appraisal and terminations: A review of court decisions since *Brito v. Zia* with implications for personnel practices. *Personnel Psychology, 40,* 489–503.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Develop-

ment of the selection procedural justice scale (SPJS). *Personnel Psychology, 54,* 389–419.

Cathcart, D. A. (1996). Employment termination litigation: Collateral tort theories and the multimillion dollar verdict: A 1994 update. In American Law Institute's *Employment Discrimination and Civil Rights Actions in Federal and State Courts* (pp. 675–779). Philadelphia: Author.

Civil Service Regulations, 5 C.F.R. § 5.1 *et seq.* (2004).

Employment Practices: Basic Requirements, 5 C.F.R. § 300.103 (2004).

Flint, D. H. (1999). The role of organizational justice in multi-source performance appraisal: Theory-based applications and directions for research. *Human Resource Management Review, 9,* 1–20.

Goldman, B. M. (2001). Toward an understanding of employment discrimination claiming: An integration of organizational justice and social information processing theories. *Personnel Psychology, 54,* 361–386.

Harvey, R. J. (1991). Job analysis. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2, pp. 71–163). Palo Alto, CA: Consulting Psychologists Press.

Klieman, A. F., & Durham, L. (1981). Performance appraisal, promotion and the courts: A critical review. *Personnel Psychology, 34,* 103–120.

Landy, F. J., & Conte, J. M. (2004). *Work in the 21st century: An introduction to industrial and organizational psychology.* New York: McGraw-Hill.

Lubben, G. L., Thompson, D. E., & Klasson, C. R. (1980, May–June). Performance appraisal: The legal implications of Title VII. *Personnel,* pp. 11–21.

Malos, S. B. (1998). Current legal issues in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 49–94). San Francisco: Jossey-Bass.

Martin, D. C., Bartol, K. M., & Kehoe, P. E. (2000). The legal ramifications of performance appraisal: The growing significance. *Public Personnel Management, 29,* 379–405.

Posthuma, R. A. (2003). Procedural due process and procedural justice in the workplace: A comparison and analysis. *Public Personnel Management, 32,* 181–195.

Roehling, M. V. (1993). "Extracting" policy from judicial opinions: The dangers of policy capturing in a field setting. *Personnel Psychology, 46,* 477–502.

Schwartz, M., & Moayed, T. (2001). Minimizing the likelihood of employment litigation. *Employee Rights Quarterly, 1,* 53–57.

Selmi, M. (2003). The price of discrimination: The nature of class action employment discrimination litigation and its effects. *Texas Law Review, 81,* 1249–1335.

Thompson, D. E., & Thompson, T. A. (1982). Court standards for job analysis in test validation. *Personnel Psychology, 35,* 865–874.

Thrasher, K. (2003). A direct defense: Using an employment practices compliance approach to avoid employee lawsuits. *Employee Relations Law Journal, 29,* 25–37.

Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 *et seq.* (1978).

Weisenfeld, D. B. (2003). The year employment law dominated the docket: How the Supreme Court's 2001–02 term affects employees. *Employee Rights Quarterly, 3,* 1–7.

Werner, J. M., & Bolino, M. C. (1997). Explaining U.S. courts of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology, 50,* 1–24.

Williams, S. D., Slonaker, W. M., & Wendt, A. C. (2003, Winter). An analysis of employment discrimination claims associated with layoffs. *SAM Advanced Management Journal,* pp. 49–55.

Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., & Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82,* 900–912.

***Cases***

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975).

Allen et al. v. City of Chicago, 351 F.3d 306 (7th Cir. 2003).

Association of Mexican-American Educators et al. v. State of California, 231 F.3d 572 (9th Cir. 2000).

Bryant v. City of Chicago, 200 F.3d 1092 (7th Cir. 2000).

Cardenas v. Massey et al., 269 F.3d 251 (3d Cir. 2001).

Cerutti et al. v. BASF et al., 349 F.3d 1055 (7th Cir. 2003).

Clady v. County of Los Angeles, 770 F.2d 1421 (9th Cir. 1985).

Cotter et al. v. City of Boston et al., 323 F.3d 160 (1st Cir. 2003).

Craig v. County of Los Angeles, 626 F.2d 659 (9th Cir. 1980).

Cullom v. Brown, 209 F.3d 1035 (7th Cir. 2000).

Dorsey v. Pinnacle Automation Co., 278 F.3d 830, 837 (8th Cir. 2002).

Espinoza v. Farah Mfg. Co., 414 U.S. 86 (1973).

Fine v. Ryan International Airlines, 305 F.3d 746 (7th Cir. 2002).

Griggs v. Duke Power Co., 401 U.S. 424 (1971).

Gulino et al. v. Board of Education, 236 F. Supp.2d 314 (S.D. N.Y. 2002).

Gulino et al. v. Board of Education, 2003 U.S. Dist. LEXIS 56 (S.D. N.Y. 2003).

Hawkins v. Pepsico, 203 F.3d 247 (4th Cir. 2000).

Haywood v. Lucent Technologies, 323 F.3d 524 (7th Cir. 2003).

Johnson v. Kroger, 319 F.3d 858 (6th Cir. 2003).
Lanning et al. v. Southeastern Pennsylvania Transportation Authority, 181 F.3d 478 (3d Cir. 1999).
Lanning et al. v. Southeastern Pennsylvania Transportation Authority, 308 F.3d 286 (3d Cir. 2002).
Liu v. Amway Corp., 347 F.3d 1125 (9th Cir. 2003).
Mayer v. Nextel, 318 F.3d 803 (8th Cir. 2003).
McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).
Meeker et al. v. Merit Systems Protection Board, 319 F.3d 1368 (Fed. Cir. 2003).
O'Neal v. Ferguson Construction Co., 237 F.3d 1248 (10th Cir. 2001).
Peele v. County Mutual Insurance Co., 288 F.3d 319 (7th Cir. 2002).
Petit et al. v. City of Chicago, 2003 U.S. App. LEXIS 25221.
Pryor v. Seyfarth, Shaw, Fairweather & Geraldson, 212 F.3d 976 (7th Cir. 2000).
Reynolds et al. v. Alabama Department of Transportation, 338 F.3d 1221 (11th Cir. 2003).
Russell v. Principi, 257 F.3d 815 (D.C. Cir. 2001).
Sauzek and Koski v. Exxon Coal USA, 202 F.3d 913 (7th Cir. 2000).
Scott v. Parkview Memorial Hospital, 175 F.3d 523, 525 (7th Cir., 1999).
Shackelford v. Deloitte & Touche, LLP, 190 F.3d 398 (5th Cir. 1999).
Shannon v. Bell South Telecommunications, Inc., 292 F.3d 712 (11th Cir. 2002).
Spears v. Department of Corrections, 210 F.3d 850 (8th Cir. 2000).
Stout et al. v. Potter, 276 F.3d 1118 (9th Cir. 2002).
Thomas v. Eastman Kodak Co., 183 F.3d 38, 51 (1st Cir. 1999).
Thorn and Curran v. Sunderstrand Aerospace Corp., 207 F.3d 383, 389 (7th Cir. 2000).
Tyler et al. v. Union Oil Co., 304 F.3d 379, 396–397 (5th Cir. 2002).
Vaughn v. Edel, 918 F.2d 517 (5th Cir. 1990).
Watson v. Fort Worth Bank & Trust, 487 U.S. 977 (1988).
Williams et al. v. Ford Motor Co., 187 F.3d 533 (6th Cir. 1999).
Winarto v. Toshiba American Electronics Components, 274 F.3d 1276 (9th Cir. 2001).
Yates v. Rexton, Inc., 267 F.3d 793 (8th Cir. 2001).