

### Establishing That Tests are Job-Related and Consistent with Business Necessity

So what makes a test defensible? A test can be successfully defended—despite its discriminatory effects—if the employer can show that the test is **job-related** and **consistent with business necessity** (or that it is a “reasonable factor other than age”). An employer has a legitimate interest in selecting people with the ability to perform jobs well. If more men than women are suited to some jobs (or vice versa), discrepant test results might reflect this fact. But employers bear the burden of proving the job relatedness of tests that create adverse impact. In *EEOC v. Dial Corp.*, the employer is unable to successfully defend its use of a physical strength test.

#### EEOC v. Dial Corp.

469 F.3d 735 (8th Cir. 2006)

##### OPINION BY CIRCUIT JUDGE MURPHY:

The Equal Employment Opportunity Commission (EEOC) brought this sex discrimination action against The Dial Corporation under Title VII of the Civil Rights Act of 1964 on behalf of a number of women who had applied for work but were not hired. \*\*\* [T]he district court concluded that Dial’s use of a preemployment strength test had an unlawful disparate impact on female applicants . . . \*\*\* We . . . affirm.

Dial is an international company with a plant located in Fort Madison, Iowa that produces canned meats. Entry level employees at the plant are assigned to the sausage packing area where workers daily lift and carry up to 18,000 pounds of sausage, walking the equivalent of four miles in the process. They are required to carry approximately 35 pounds of sausage at a time and must lift and load the sausage to heights between 30 and 60 inches above the floor. Employees who worked in the sausage packing area experienced a disproportionate number of injuries as compared to the rest of the workers in the plant.

Dial implemented several measures to reduce the injury rate starting in late 1996. These included an ergonomic job rotation, institution of a team approach, lowering the height of machines to decrease lifting pressure for the employees, and conducting periodic safety audits. In 2000 Dial also instituted a strength test used to evaluate potential employees, called the Work Tolerance Screen (WTS). In this test job applicants were asked to carry a 35 pound bar between two frames, approximately 30 and 60 inches off the floor, and to lift and load the bar onto these frames. The applicants were told to work at their “own pace” for seven minutes.

An occupational therapist watched the process, documented how many lifts each applicant completed, and recorded her own comments about each candidate’s performance. Starting in 2001, the plant nurse, Martha Lutenegger, also watched and documented the process. From the inception of the test, Lutenegger reviewed the test forms and had the ultimate hiring authority.

For many years women and men had worked together in the sausage packing area doing the same job. Forty six percent of the new hires were women in the three years before the WTS was introduced, but the number of women hires dropped to fifteen percent after the test was implemented. During this time period the test was the only change in the company’s hiring practices. The percentage of women who passed the test decreased almost each year the test was given, with only eight percent of the women applicants passing in 2002. The overall percentage of women who passed was thirty eight percent while the men’s passage rate was ninety seven percent. While overall injuries and strength related injuries among sausage workers declined consistently after 2000 when the test was implemented, the downward trend in injuries had begun in 1998 after the company had instituted measures to reduce injuries. \*\*\*

A jury trial was held in August 2004, and EEOC and Dial offered testimony by competing experts. EEOC presented an expert on industrial organization who testified that the WTS was significantly more difficult than the actual job workers performed at the plant. He explained that although workers did 1.25 lifts per minute on average and rested between lifts, applicants who took the WTS performed 6 lifts per minute on average, usually without any breaks. He also testified that in two

of the three years before Dial had implemented the WTS, the women's injury rate had been lower than that of the male workers. \* \* \*

Dial presented an expert in work physiology, who testified that in his opinion the WTS effectively tested skills which were representative of the actual job, and an industrial and organizational psychologist, who testified that the WTS measured the requirements of the job and that the decrease in injuries could be attributed to the test. Dial also called plant nurse Martha Lutenecker who testified that although she and other Dial managers knew the WTS was screening out more women than men, the decrease in injuries warranted its continued use. \* \* \*

The district court . . . found that the WTS had had a discriminatory effect, that Dial had not demonstrated that the WTS was a business necessity or shown either content or criterion validity, and that Dial had not effectively controlled for other variables which may have caused the decline in injuries, including other safety measures that Dial had implemented starting in 1996. \* \* \*

Statistical disparities are significant if the difference between the expected number and the observed number is greater than two or three standard deviations. Here, the disparity between hiring of men and women showed nearly ten standard deviations. The percentage of women who passed the WTS declined with each implementation of the test. Despite knowing about the statistical difference, Dial continued to use the WTS. Dial argues that EEOC's statistics are inapplicable because men and women are not similarly situated and have profound physiological differences. There was evidence, however, that women and men worked the same job together for many years before the WTS was instituted. \* \* \*

In a disparate impact case, once the plaintiff establishes a prima facie case the employer must show the practice at issue is "related to safe and efficient job performance and is consistent with business necessity." An employer using the business necessity defense must prove that the practice was related to the specific job and the required skills and physical requirements of the position. Although a validity study of an employment test can be sufficient to prove business necessity, it is not necessary if the employer demonstrates the procedure is sufficiently related to safe and efficient job performance. If the employer demonstrates business necessity, the plaintiff can still prevail by showing there is a less discriminatory alternative.

Dial contends the WTS was shown by its experts to have both content and criterion validity. Under EEOC guidelines, "A content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated." Dial's physiology expert testified that the WTS was highly representative of the actions required by the job, and Dial claims that his testimony was not rebutted by EEOC, which had no physiology witness. The district court was persuaded by EEOC's expert in industrial organization and his testimony "that a crucial aspect of the WTS is more difficult than the sausage making jobs themselves" and that the average applicant had to perform four times as many lifts as current employees and had no rest breaks. There was also evidence that in a testing environment where hiring is contingent upon test performance, applicants tend to work as fast as possible during the test in order to outperform the competition.

Dial argues the WTS was criterion valid because both overall injuries and strength related injuries decreased dramatically following the implementation of the WTS. The EEOC guidelines establish that criterion validity can be shown by "empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance." Although Dial claims that the decrease in injuries shows that the WTS enabled it to predict which applicants could safely handle the strenuous nature of the work, the sausage plant injuries started decreasing before the WTS was implemented. Moreover, the injury rate for women employees was lower than that for men in two of the three years before Dial implemented the WTS. The evidence did not require the district court to find that the decrease in injuries resulted from the implementation of the WTS instead of the other safety mechanisms Dial started to put in place in 1996. \* \* \*

Since Dial failed to demonstrate that the WTS was a business necessity, however, EEOC never was required to show the absence of a nondiscriminatory alternative. Part of the employer's burden to establish business necessity is to demonstrate the need for the challenged procedure, and the court found that Dial had not shown that its other safety measures "could not produce the same results." We conclude that the district court findings in its disparate impact analysis were not clearly erroneous, and we see no legal error in its conclusions on liability. \* \* \*

**CASE QUESTIONS**

- |  |  |
|--|--|
| <ol style="list-style-type: none"> <li>1. What was the legal issue in this case? What did the court decide?</li> <li>2. What is the evidence that use of the strength test disadvantaged women?</li> <li>3. What is content validity? What is criterion validity?</li> </ol> | <p>How did the employer attempt to show the content validity of the strength test? To show the test's criterion validity? Why was the court not convinced?</p> <ol style="list-style-type: none"> <li>4. What should Dial do at this point?</li> <li>5. Do you agree with the court's decision? Why or why not?</li> </ol> |
|--|--|

It is incumbent on employers to have evidence of the **validity** of a test that is producing discriminatory effects. In the most general terms, the validity of a measurement device or test refers to whether it actually measures what it purports to measure. Because employment tests are designed to help employers select people who will perform their jobs well, the validity of an employment test rests on how well it predicts future job performance. Does the test help employers make good selection decisions by measuring job-related criteria, or does it impose a needless obstacle?

There are two primary ways to establish the validity of employment tests. The first is **content validation**. A test has content validity to the extent that it requires the performance of the same behaviors and skills as the job in question. A classic example of a content-valid testing procedure is use of a word-processing test (e.g., speed, accuracy) to assess candidates for a secretarial job. The test is clearly representative of behaviors and skills that are central to the job. A legal limitation on content validation is that it cannot be used to validate tests of intelligence, personality, or other intangible traits. However, tests of job-specific knowledge or skills can be validated in this manner. Another problem is that the tests assume that candidates already possess and can display the desired skills. But many jobs can be learned relatively quickly. Testing procedures that screen out people who would be capable of performing the job following minimal amounts of on-the-job training and practice are not job-related and consistent with business necessity. *Thus, employers should use content-valid tests as much as possible but not to assess intangible characteristics such as intelligence and not for tasks that could readily be learned on the job.*

The city of Buffalo, New York, was able to defend its fire department's exam for promotion to fire lieutenant against an adverse impact claim brought by African American firefighters by showing the test's content validity.<sup>49</sup> To establish the test's content validity, the court said that it was necessary for the city to prove that (1) the test makers conducted a suitable job analysis, (2) the test makers used reasonable competence in constructing the test itself, (3) the content of the test related to the content of the job, (4) the content of the test was representative of the content of the job, and (5) the test scoring system usefully selected from among the applicants those who can better perform the job.<sup>50</sup> Although the court faulted the city for relying on a statewide job analysis rather than carefully studying the fire lieutenant job in its own department—and failure to conduct a thorough examination of the tasks and responsibilities of the specific job in question is usually a fatal flaw in a validation study—there was sufficient evidence in this case that the job of fire lieutenant varied little across cities in the state. Reasonable competence in test preparation was shown by the hiring of professional test preparers and use of questions that had been previously assessed for comprehensibility. The court was satisfied that the test preparer had identified key aspects of the fire lieutenant job and developed questions designed to tap relevant knowledge. In contrast, courts have ruled against the New York City Board of Education

<sup>49</sup> *M.O.C.H.A. Society v. City of Buffalo*, 689 F.3d 263 (2d Cir. 2012).

<sup>50</sup> *M.O.C.H.A. Society*, at 274.