

In the
United States Court of Appeals
For the Seventh Circuit

No. 22-2433

CATHERINE ERDMAN,

Plaintiff-Appellant,

v.

CITY OF MADISON,

Defendant-Appellee.

Appeal from the United States District Court for the
Western District of Wisconsin.
No. 3:16-cv-00786-wmc — **William M. Conley**, *Judge*.

ARGUED SEPTEMBER 14, 2023 — DECIDED JANUARY 22, 2024

Before ROVNER, HAMILTON, and BRENNAN, *Circuit Judges*.

HAMILTON, *Circuit Judge*. Title VII of the Civil Rights Act of 1964 prohibits sex discrimination in hiring, but it allows the use of selection criteria that are tied sufficiently to the actual job demands even if those criteria may have disparate impacts on women and men. See 42 U.S.C. § 2000e-2(h) & (k). Fighting fires is one of the most physically demanding jobs around. Tests of physical abilities for firefighters have been the subject of both study and Title VII litigation. This case is an example.

Plaintiff Catherine Erdman applied to work as a firefighter for the city of Madison, Wisconsin. She was eliminated from the fire department's 2014 recruitment class after failing to achieve a minimum qualifying score in the final event of the department's Physical Abilities Test (which we call "the Madison test"). Erdman filed this civil action against the city claiming that the Madison test has a disparate impact on women in violation of Title VII. After a bench trial, the district court found that Erdman had failed to prove her claim that the department's physical abilities test violated Title VII. *Erdman v. City of Madison*, 615 F. Supp. 3d 889 (W.D. Wis. 2022).

The district court found that Erdman had shown the Madison test had a prima facie disparate impact on women. The court also found, however, that the Madison test was job-related and served the city's legitimate needs. The court also found that Erdman had failed to prove that her proposed alternative hiring practice would serve the city's legitimate needs. The use of the Madison physical abilities test to disqualify Erdman thus did not violate Title VII. We affirm.

I. *Factual And Procedural Background*

A. *The Madison Fire Department's Hiring*

The Madison fire department employs approximately 365 firefighters. The department uses a formal recruitment process to hire new firefighters. The hiring process consists of five stages: (1) an application screening for minimum qualifications; (2) a written test; (3) a physical abilities test; (4) an oral examination before a panel of examiners; and (5) an interview with the chief.

Plaintiff Catherine Erdman has been a firefighter for the city of Janesville, Wisconsin, since 2007. She applied for a

position with the Madison fire department in the 2014 hiring process but did not pass the Madison physical abilities test. She alleges that the Madison physical abilities test caused the city to discriminate against her on the basis of sex.

A total of 1,887 applicants participated in the 2014 hiring process. Of these, 1,723 were men, 146 were women, and 18 were not clearly identified in records by gender. Four hundred and ninety-nine applicants appeared to take the Madison physical ability test — 471 men and 28 women. Of these, 404 applicants — 395 men, four women, and five not clearly identified — successfully completed the test. Ultimately that year, the department hired all four women who passed the Madison test, as well as thirteen men. Overall in 2014, 14 % of Madison firefighters were women. In 2018, 10.8% were women. Both numbers were well above the national average of about 4% female firefighters.

1. *The 2014 Madison Physical Abilities Test*

The posting for Madison's 2014 firefighter recruitment identified the following physical requirements for working as a firefighter:

While not an exclusive list, the following examples are meant to illustrate some of the extreme physical demands and working conditions inherent in the role of a firefighter.

Physical Demands

1. Pick up and advance charged fire hoses.*
2. Force entry with axe/battering ram.*
3. Rescue/extricate victim(s).*
4. Perform CPR; apply bandages.

5. Climb stairs carrying heavy equipment, while wearing firefighter protective clothing that weighs in excess of 50 pounds.*
6. Strip and vent roofs, breach walls, overhaul burned buildings.*
7. Lift and climb/descend ladders (with victims).*
8. Visually determine fire status/hazards; assess patient conditions.
9. Hear calls for help; identify fire noise, etc.
10. Walk on roof tops under adverse conditions.
11. Operate power tools and extrication equipment; tie knots.
12. Stoop, crawl, crouch, and kneel in confined spaces.*
13. Reach, twist, balance, grapple, bend and lift under emergency conditions.
14. Run, dodge, jump and maneuver with equipment.*
15. All of the above may be performed wearing heavy and restrictive protective clothing/gear in excess of 50 pounds.*

Each task marked by an asterisk was assessed in the 2014 Madison test. That test consisted of seven events: (1) equipment shuttle; (2) ladder event; (3) hose drag; (4) sledgehammer event; (5) search; (6) rescue; and (7) pike pole.

The seven events had to be completed in order, but they were timed and scored separately. To explain the issues on

appeal, we must explain scoring for the test. Each event was assigned both a so-called “cut score” and a lower “minimally acceptable score.” To pass the test, the applicant had to achieve the higher “cut score” in at least five of the seven events. To pass the entire test, the applicant also had to achieve at least the minimally acceptable score in each of the seven events.

2. Erdman’s Application History

Erdman entered firefighting first as a volunteer in Poynette, Wisconsin, and then, in 2007, as a full-time firefighter in Janesville, Wisconsin. At Janesville, in a 90-person fire department, she was promoted several times. After being nominated by her peers and chosen by the fire chief, she received the Janesville Firefighter of the Year award in 2014. At the time of the trial in 2018, she had been deployed to about 230 fires, about 60 to 65 of which were structure fires.

Erdman has applied for a position with the Madison fire department in every hiring cycle since 2006. She failed the written examination on five occasions, in 2006, 2008, 2010, 2012, and 2018. During the 2014 cycle, Erdman passed the written examination and proceeded to the next step, the Madison physical abilities test.

Erdman met the cut score and received points for five of the seven events: equipment shuttle, hose drag, sledgehammer event, search, and rescue. While Erdman did not meet the cut score for the ladder event, she nevertheless attained the minimum acceptable score required to avoid disqualification. Whether she passed or failed the entire test came down to the final event, the pike pole test. In this event, the applicant must use a pole with a hook on it first to simulate breaching a

ceiling from below to look for hidden flames, and then pulling down ceiling material.

The minimum acceptable score for that event was 16 repetitions in the time allowed. Erdman completed only 12 repetitions. That score eliminated her from the 2014 hiring process. If she had met the pike pole event's minimally acceptable score of 16, she would have passed the entire test and moved on to the next stage in the hiring process.

3. *The Alternative: The IAFF Candidate Physical Abilities Test*

Erdman contends that a different physical abilities test would have had less disparate impact on female applicants but would have sufficiently served the city's purpose in testing applicants' physical abilities to work as firefighters. As her alternative, Erdman proposes the Candidate Physical Abilities Test. It is licensed by the International Association of Fire Fighters (IAFF) and used as a screening tool for many fire departments across the nation. The test was developed in conjunction with ten fire departments in North America, including New York City, Indianapolis, and Austin, Texas.

The district court heard detailed evidence on the similarities and differences between the Madison test and the IAFF test. The IAFF test contains eight component parts, seven of which are identical or similar to events in the Madison test.

Differences include the method of timing each test. The IAFF test requires applicants to complete all events within a designated total time but does not put a time limit on individual events. The Madison test times each event separately. The IAFF test also allows candidates several chances to pass the test. A candidate may take the test up to three times, and he

or she must pass only once. In the Madison recruiting process, each applicant has only one chance to pass the Madison physical ability test.¹

Another difference involves the pike pole event that disqualified Erdman in 2014. In the IAFF test, as an applicant performs the “pull” portion of the test, she is allowed to stand directly under the point where she has hooked her pole. In the Madison test, applicants can stand directly under the hook point to push up, but they must move 18 inches back when completing the “pull” portion of the test.

Since at least 2013, the Madison fire department has been aware of the IAFF test, considered it, and decided to continue using its own test. The district court found that the IAFF test has less of a disparate impact on women than the Madison test. On appeal, the city does not challenge that finding, at least as applied to the entire test. But the district court agreed with the city that the IAFF test would not adequately serve the legitimate needs of Madison’s fire department. For instance, the city argued that if the department had offered the

¹ As a result of a 2006 conciliation agreement between the Equal Employment Opportunity Commission and the International Association of Firefighters, takers of the IAFF test are also afforded practice sessions in the weeks before the test that offer candidates hands-on experience with testing equipment under the guidance of trainers who coach candidates on how to successfully complete the component tasks of the test. To the extent that one’s success with a task-oriented physical abilities test like the Madison test or IAFF test depends on technique as well as physical strength, these practice sessions provide a concrete benefit to any candidate, male or female. But Erdman’s expert witness, Professor Arthur Weltman, was unaware of any study documenting that female candidates benefit more from pre-testing practice and training than do their male counterparts. R. 67 at 27–28.

IAFF test's required training to the nearly 500 applicants who performed the Madison test in 2014, it would have resulted in significant cost and overtime expense. The district court also agreed with the city that, unlike the IAFF test, "certain elements of the [Madison test] were designed specifically for Madison, in light of characteristics of the city, the Department's equipment or other considerations, including safety."

B. *Procedural History*

After discovery, both sides moved for summary judgment. The district court granted summary judgment to the city on one issue raised on appeal, Erdman's ability to recover back pay and front pay, but we need not reach that issue. The district court denied summary judgment on the remaining issues presented in this appeal, and the parties proceeded to trial.

After a bench trial, the district court found that Erdman had failed to prove that the Madison test had a disparate impact on female applicants in violation of Title VII. The court found that the Madison test created a prima facie disparate impact on women. Erdman's claim failed, however, because she did not show that her proposed alternative hiring practice, the IAFF test, would serve the department's legitimate needs as well as the Madison test.

II. *Discussion*

After a district court makes findings of fact and conclusions of law after a bench trial, we review legal conclusions de novo and factual findings for clear error. Fed. R. Civ. P. 52(a); *Simon v. Cooperative Educational Service Agency #5*, 46 F.4th 602, 606 (7th Cir. 2022), quoting *Murdock & Sons Construction, Inc. v. Goheen General Construction, Inc.*, 461 F.3d 837, 840 (7th Cir.

2006). On clear-error review, a finding of fact is reversed “only when the reviewing court on the entire evidence is left with the definite and firm conviction that a mistake has been committed.” *Coleman v. Comm’r of Internal Revenue*, 16 F.3d 821, 826 (7th Cir. 1994), quoting *Anderson v. City of Bessemer City*, 470 U.S. 564, 573 (1985). For mixed questions of law and fact, our standard of review depends on whether the inquiry “entails primarily legal or factual work.” *U.S. Bank N.A. v. Village at Lakeridge, LLC*, 583 U.S. 387, 396 (2018). We discuss the applicable standards of review in more detail as we apply them to the district court’s findings and conclusions and the parties’ arguments on appeal.

A. *Title VII’s Burden-Shifting Framework for Disparate Impact in Hiring Practices*

Title VII prohibits hiring practices that have a disproportionate adverse impact on job applicants with protected characteristics such as sex or race, even in the absence of discriminatory intent. *Ernst v. City of Chicago*, 837 F.3d 788, 794 (7th Cir. 2016) (speaking in terms of “employees” but also applying statute to applicants); see 42 U.S.C. § 2000e–2(a)(2); *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) (“individual” applies to applicants who assert disparate impact in hiring).

To prove disparate impact, a plaintiff must show that a particular hiring practice had an adverse impact on applicants with a protected characteristic, such as sex. § 2000e–2(k)(1)(A)(i); *Ernst*, 837 F.3d at 796. As part of the prima facie case, the plaintiff-applicant must also show that the challenged hiring practice causes the discriminatory impact, typically by offering “statistical evidence ... sufficient to show that the practice in question has caused the exclusion of applicants ... because of their membership in a protected

group.” *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 994 (1988).

If a job applicant makes a prima facie showing of disparate impact, an employer can defend by showing that: (1) the challenged practice does not cause the disparate impact, § 2000e–2(k)(B)(ii); or (2) the practice is job-related for the position and consistent with business necessity. § 2000e–2(k)(1)(A)(i); see generally *Griggs*, 401 U.S. at 431 (explaining disparate impact test). If the employer can show that the practice is job-related for the position and consistent with business necessity, the burden shifts back to the applicant to prove that the employer refuses to adopt an alternative hiring practice that would result in less disparate impact and still serve the employer’s legitimate needs. § 2000e–2(k)(1)(A)(ii) & (C); *Ernst*, 837 F.3d at 794.

Erdman relies on this three-step burden-shifting framework. After a bench trial, the district court found at step one that Erdman had established her prima facie case that the Madison test has a disparate impact on women. The court found at the second step that the city showed the Madison test was job-related and consistent with business necessity. At the third step, Erdman needed to prove that the city refused to adopt an alternative hiring practice (in this case, the IAFF test) that would have less disparate impact and still serve the city’s legitimate needs. The district court found that while the IAFF test had less disparate impact on women, Erdman failed to prove that it would serve the city’s legitimate needs as well as the Madison test.

On appeal, the city contests the district court’s finding that Erdman proved her prima facie case of disparate impact. Erdman challenges the district court’s finding that the IAFF test

fails to serve adequately the fire department's legitimate needs. We consider each issue in turn.

B. *Prima Facie Case of Disparate Impact*

We begin at the first step, Erdman's prima facie case of disparate impact. Although we affirm at the third step, we start with the first because all three steps of Title VII's burden-shifting framework are logically intertwined. The parties dispute whether the proper unit of analysis for the prima facie disparate impact is the entire Madison test or just the particular event that disqualified Erdman, the pike pole event. The answer determines the scope of the practice the city had to justify at the second step and the alternative hiring practice Erdman needed to offer at the third step.

At step one in a disparate impact case, the applicant must show that the employer "uses a particular employment practice that causes a disparate impact" on an impermissible basis. § 2000e-2(k)(1)(A)(i); see *Ernst*, 837 F.3d at 796. Erdman asserts that the Madison test in its entirety has an illegal disparate impact on female applicants. The city argues that the analysis must focus more narrowly, on only the specific part of the Madison test that disqualified plaintiff. Of the seven women who reached the pike pole event in the 2014 Madison test, six passed; only Erdman was disqualified. Seeing no proof of disparate impact specific to just the pike pole event, the city argues that Erdman has not made her prima facie showing.

The city concedes on appeal that the Madison test as a whole shows a statistically significant disparate impact on female applicants. In 2014, the pass rate for women who appeared to take the test was 14% (4 out of 28), while the pass rate for men who appeared to take the test was 84% (395/471).

The disqualification rate for women who appeared to take the test and did not quit (20/25) was more than seven times the disqualification rate of men who appeared to take the test and did not quit (49/458). Whether the “particular employment practice” at issue is the entire Madison physical abilities test or just the pike pole event will resolve the issue of Erdman’s prima facie case. The district court held that the entire Madison test was the proper unit of analysis. We agree.

1. *The City’s Procedural Problem*

On appeal, the city has framed this issue as a challenge to the district court’s denial of summary judgment, not the findings after trial. That’s a problem. Denial of summary judgment “is an interlocutory matter subsumed by a final judgment. ... After trial, the summary judgment denial is ancient history and not subject to appeal.” *Empress Casino Joliet Corp. v. Balmoral Racing Club, Inc.*, 831 F.3d 815, 823–24 (7th Cir. 2016). The Supreme Court has recently recognized an exception for pure issues of law, *Dupree v. Younger*, 598 U.S. 729, 735–36 (2023), but the prima facie case presents a mixed question of law and fact. We can review a denial of summary judgment on a mixed question only if the party appealing the denial filed motions for judgment as a matter of law and renewed judgment as a matter of law after trial. *Empress Casino*, 831 F.3d at 823, citing *Ortiz v. Jordan*, 562 U.S. 180, 189–90 (2011).

Here, the city did not file either motion, abandoning its path to challenge as of right the district court’s final judgment on this prima facie issue. Given the lack of any reference to *Dupree* or the final-order doctrine in the city’s brief, this seems to have been at least initially an inadvertent forfeiture by the city. At oral argument, the city defended its choice to

challenge the district court's denial of summary judgment under *Dupree* as an appeal of a pure legal issue, making the city's actions seem more akin to waiver.

Whether the issue was waived or forfeited, however, a feature of this case leads us to reach the merits of the challenge. As we recently explained, “[w]hether a waived issue can be addressed ‘is one left primarily to the discretion of the courts of appeals, to be exercised on the facts of individual cases.’” *Walker v. Baldwin*, 74 F.4th 878, 883 (7th Cir. 2023), quoting *Singleton v. Wulff*, 428 U.S. 106, 121 (1976). “[T]he refusal to consider arguments not raised is a sound prudential practice, rather than a statutory or constitutional mandate, and there are times when prudence dictates the contrary.” *Davis v. United States*, 512 U.S. 452, 464 (1994) (Scalia, J., concurring). Though such exercises of discretion are “rare,” see *Mahran v. Advocate Christ Medical Center*, 12 F.4th 708, 713 (7th Cir. 2021), in this case it would be most prudent to address the merits of the city's argument on the scope of a “particular employment practice.”

Doing so here does not contradict the purposes of the party presentation principle. That principle “ensures that the opposing party has an opportunity to reflect upon and respond in writing to the arguments that his adversary is raising.” *Berkman v. Vanihel*, 33 F.4th 937, 947 n.45 (7th Cir. 2022). Here, the *prima facie* case was fully litigated in the district court, and Erdman could and did address it adequately (and persuasively) on appeal in her reply brief. Most important, the proper scope of a “particular employment practice” is entwined with plaintiff's burden to prove an adequate “alternative hiring practice” at the last step of Title VII's burden-shifting framework. It would be

burdensome for these and similarly situated parties to require them to litigate business justifications and alternative hiring practices when the proper scope of the case remains in such doubt. We exercise our discretion to reach the merits of the prima facie case.

2. *The Prima Facie Case for a Disparate Impact*

The prima facie case presents a mixed question of law and fact. Whether the Madison test as a whole or its components constitute a “particular employment practice” requires the courts to “interpret[] the legal significance of a unique series” of physical ability testing events. See § 2000e–2(k)(1)(A)(i); *Straits Financial LLC v. Ten Sleep Cattle Co.*, 900 F.3d 359, 368 (7th Cir. 2018) (review of mixed questions of law and fact). The district court’s conclusion that the proper unit of analysis was the Madison test as a whole was tied closely to the facts of this case. We apply clear-error review to the case-specific conclusions while also “taking care not to defer to the district court’s judgment where it effectively announced new legal rules that would govern future cases.” *Straits Financial*, 900 F.3d at 368.

We look first to the district court’s analysis of precedent on the meaning of a “particular employment practice” in 42 U.S.C. § 2000e–2(k)(1)(A)(i). Because adopting this portion of the district court’s reasoning would govern future cases and because our review entails primarily legal work rather than factual work, we review de novo. See *Village at Lakeridge*, 583 U.S. at 396. We agree with the district court that nothing in *Watson* or any other binding case law requires a plaintiff to isolate a particular component of a physical ability test or similar multi-component test as a “particular employment practice” for a disparate impact claim. See *Watson*, 487 U.S. at

994 (“identifying the specific employment practice ... has been relatively easy to do in challenges to standardized tests”).

As the district court explained, the city’s event-by-event analysis is not supported by precedent:

[N]either the parties nor this court in its own research could find a case analyzing whether a “particular employment practice” must be limited to the particular component of a physical abilities test or of a similar multi-component test. [Other] courts evaluating similar cases involving [physical ability tests] appear to have uniformly considered the entire [test] as an indivisible hiring practice. *See, e.g., Pietras v. Bd. of Fire Comm’rs of Farmingville Fire Dist.*, 180 F.3d 468, 475 (2d Cir. 1999) (comparing pass rate for women and pass rate for men on [physical abilities test] as a whole); *Arndt v. City of Colorado Springs*, 263 F. Supp. 3d 1071, 1075 (D. Colo. 2017) (requirement that all sworn officers pass [physical abilities test] annually was a specific employment practice); *cf. Ernst*, 837 F.3d at 796 (“[P]hysical-skills entrance test has an adverse impact on women[.]”).

Erdman v. City of Madison, No. 16-cv-786-wmc, 2018 WL 4496308, at *6 (W.D. Wis. Sept. 19, 2018). The parties have not cited such a case to us, either, nor have we found one in our own research.

The city argues that Title VII’s text forbids plaintiffs from focusing on a broader decision-making process if that process

is composed of discrete parts that are capable of separation. The city relies on language in § 2000e-2(k)(1)(B)(i): “if the complaining party can demonstrate to the court that the elements of a respondent’s decisionmaking process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice.” See also *Davis v. Cintas Corp.*, 717 F.3d 476, 497 (6th Cir. 2013) (plaintiff must show that hiring system’s “many steps were so intertwined that they were not capable of separation for analysis”).

The statute’s text tells us that separability is plaintiff’s burden, but the text does not specify features that render a decision-making process, in the language of § 2000e-2(k)(1)(B)(i), “capable of separation.” *Davis* does not help the city because the plaintiff in *Davis* “did not explain why the well-defined, discrete elements” of that hiring system were “not capable of separation for analysis.” 717 F.3d at 497. Because the plaintiff there failed to carry even her burden of production on separability, the court had no occasion to state or apply a test for when multiple steps in a hiring process were “so intertwined that they were not capable of separation for analysis.” *Id.*

But here, as the district court noted, Erdman raised strong legal and factual arguments to support analyzing the entire Madison test as a singular “hiring practice.” Because the question of separability is fact-bound, we review the district court’s analysis of separability for clear error. We agree with the district court that plaintiff demonstrated that the Madison test is not separable for analysis into its discrete elements.

On appeal, the city leans heavily on one district court case to show that courts can “successfully analyze[]” a single component of a multi-part physical fitness test. See *Easterling v.*

State of Connecticut, 783 F. Supp. 2d. 323 (D. Conn. 2011). The physical fitness test challenged in *Easterling* consisted of four separately timed and scored components: a sit-and-reach test, a one-minute sit-up test, a one-minute push-up test, and a timed 1.5-mile run. 783 F. Supp. 2d. at 326. Based on *Easterling*, the city urges us to adopt as a rule that if subcomponents of a particular multi-component test are timed or scored separately, they are capable of separation and must be analyzed individually for disparate impact.

But as Erdman notes, in *Easterling*, the plaintiff *chose* to challenge only the timed 1.5-mile-run element of a physical fitness test. In that case, neither party even suggested that the analysis ought to be conducted at the level of the whole test. The issue here—whether to separate the physical fitness test’s different events—was not presented to the court, nor did the court discuss it.

Even if we treated *Easterling* as a meaningful holding on the separability of the multi-part test at issue in that case, that fitness test was distinct from the Madison test in ways that weigh against separability here. The employer-defendant in *Easterling* said that each element of its test measured distinct, non-overlapping physical abilities: respectively, (1) “flexibility of the lower back and upper leg area,” (2) “muscular endurance of the abdominal muscles,” (3) “muscular endurance of the chest, upper arms and shoulders (upper body dynamic strength),” and (4) “the heart and vascular system’s capability to transport oxygen (cardiovascular endurance).”

The Madison test, conversely, “focuses on job simulation activities, rather than traditional exercise-based activities.” That is, the Madison test does not test distinct physical capabilities, but instead evaluates the sorts of activities firefighters

might reasonably be expected to do on the job. Accordingly, it requires participants to use the same muscle groups and elements of physical fitness repeatedly in a short time. One of plaintiff's witnesses here, Madison Assistant Chief Popovich, testified that by the time applicants reach the pike pole test, "the seventh of seven stations," they are usually "very, very exhausted." That evidence also helped persuade the district court to deny summary judgment on the issue: the "combined demands of all of the components of the [Madison test], which are extremely physical and occur consecutively," supported treating the test as a single hiring practice. *Erdman*, 2018 WL 4496308, *7. After trial the district court stood by this reasoning. 615 F. Supp. 3d at 896. It has ample support in the record.

The city's proposed standard for separating multi-part hiring tests strikes us as unduly simplistic. It would also tend to undermine Title VII's protections from disparate impacts of hiring tests that are not justified in terms of business necessity. Judge Conley recognized these dangers: "if multiple successive components of the PAT are collectively discriminatory, as plaintiff claims they are, she need not demonstrate a disparate impact as to the last component simply because she succeeded in overcoming earlier discriminatory components." 2018 WL 4496308, *7. He added, "if an employer uses a hiring practice that discriminatorily eliminates members of the protected class at each event level, the employer may not be shielded from liability because of the smaller sample size at each successive event level." *Id.*

In a multi-part test, it may be that each individual component eliminates only a small fraction of a minority group, but with a cumulative effect that excludes a significant proportion of minority applicants from employment. The city's proposed

standard for separability would encourage, or at least allow, gamesmanship in structuring multi-element tests to limit the set of plaintiffs who could bring discrimination challenges. In cases like this one, where poor performance on one subcomponent can eliminate a candidate before she completes the entire test, plaintiffs will inevitably face smaller sample sizes at later stages of the test. Requiring plaintiffs to challenge only the particular sub-component at which they failed could make it nearly impossible for those challenging late-stage events to meet their burden of proof even where the overall effect of the test imposes a heavy disparate impact on one group compared to another.

The limits of the city's proposed rule are also unclear. Each question on a written test, for example, is typically scored separately, and, as with the Madison test of physical ability, aggregated to produce a total test score. Under the city's logic, plaintiffs could be required to bring a disparate impact challenge only to particular questions they answered incorrectly. Yet we have often treated an entire written examination as a "particular employment practice." See *Adams v. City of Chicago*, 469 F.3d 609, 610–11, 613 (7th Cir. 2006) (entire written qualifying examination is "particular employment practice"); *Allen v. City of Chicago*, 351 F.3d 306, 309, 312 (7th Cir. 2003) (same); see also *Pouncy v. Prudential Ins. Co. of America*, 668 F.2d 795, 800–01 (5th Cir. 1982) (explaining why "disparate impact analysis may be used to challenge aptitude and intelligence tests" but not "entire range of employment practices").

The district court did not plainly err in holding that the Madison test's events are not separable for analysis, and consequently, that the proper unit of analysis here was the test as

a whole. Multi-part physical tests like the Madison fire department's test can be challenged as a "particular employment practice." They should not be deemed "capable of separation" within the meaning of Title VII merely because their events are scored and timed separately. See § 2000e-2(k)(1)(B)(i).

The city acknowledges that if the correct unit of analysis is the entire test, Erdman established her prima facie case of disparate impact. We affirm the district court's holding that plaintiff proved her prima facie case of disparate impact.

C. Less Discriminatory Alternative Employment Practice

Erdman concedes that the city carried its second-step burden to prove that the Madison test is "job-related for the position and consistent with business necessity." On appeal, she challenges only the district court's adverse third-step finding that the IAFF test is not an "available alternative employment practice that has less disparate impact and serves the employer's legitimate needs." *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009), citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii) and (C).

The city, for its part, chose not to dispute the district court's finding that the IAFF test has less disparate impact on female applicants than the Madison test. And at trial, the city did not dispute the general validity of the IAFF test. The city built its defense on the theory that the IAFF test does not serve the department's unique legitimate needs—that is, the IAFF test is not *locally* valid for the Madison fire department.

After trial, the district court found that the IAFF test would not adequately serve the fire department's legitimate needs. Because this is a factual question, we review for clear error. *Simon v. Cooperative Education Service Agency #5*, 46 F.4th

602, 606 (7th Cir. 2022). The district court did not clearly err in the inferences it made based on the record at trial.

To serve an employer's "legitimate needs," an alternative hiring practice need not be exactly as effective as the allegedly discriminatory practice. The issue is whether the alternative practice is "substantially equally valid." *Allen v. City of Chicago*, 351 F.3d 306, 312 (7th Cir. 2003), quoting 29 C.F.R. § 1607.3(B). A test is substantially equally valid if it "would also serve the employer's legitimate interest in efficient and trustworthy workmanship." *Allen*, 351 F.3d at 312, quoting *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975). Under this standard, an alternative test is "substantially equally valid" if it "would lead to a workforce [that is] substantially equally qualified." *Allen*, 351 F.3d at 314.

Allen established what amounts to a minimum-competency test. It is met if the plaintiff can establish by a preponderance of the evidence that the last person selected by the alternative approach would be roughly as qualified as the candidate with the lowest passing score using the current hiring practice. *Allen*, 351 F.3d at 313 ("The officers effectively bear the burden of establishing that the last officer promoted under an increased merit-based selection process would be roughly as qualified as the officer with the current lowest score on the assessment exercise."); see also *Ernst v. City of Chicago*, 837 F.3d at 804 ("[A] discriminatory cutoff score on an entry level employment examination must be shown to measure the minimum qualifications necessary for successful performance of the job."), quoting *Lanning v. Southeastern Pennsylvania Transp. Auth.*, 308 F.3d 286, 287 (3d Cir. 2002).

The relative effectiveness of the challenged practice and the proposed alternative in selecting a minimally qualified

workforce is not the exclusive consideration for whether two tests are “substantially equally valid.” *Allen*, 351 F.3d at 312. Factors such as the cost or other burdens of proposed alternative selection methods are also relevant in determining whether they would be substantially as efficient as the challenged practice in serving the employer's legitimate business goals. *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 998 (1988). The Supreme Court has not offered more specific guidance for such comparisons. We consider the balancing of various burdens associated with an alternative hiring practice as a mixed question of fact and law, subject to the standard of review for mixed questions described above.

The district court’s factual findings on the local validity of the IAFF test for the Madison fire department were not clearly erroneous. As the district court noted, “plaintiff bears the burden to prove the [IAFF test] would serve the Madison Fire Department’s legitimate needs.” 615 F. Supp. 3d at 900. The district court found that Erdman did not put forth sufficient affirmative evidence on this point. She “simply points to the [IAFF test], assuming that it would fit Madison’s needs without attempting to validate the test locally.” *Id.* at 899.

Validating the test locally would require at least some affirmative evidence from Erdman showing that IAFF-test candidates with the lowest passing scores would be “roughly as qualified” as those who obtain the lowest passing score on the Madison test. But to establish this point, Erdman points primarily to trial testimony by a single defense expert who was unable to give a reason why the IAFF test may not be locally valid for the Madison fire department. One expert’s inability to answer that question is not affirmative evidence to

meet Erdman's burden. It certainly does not mean the whole record would require a finding that she met her burden.

Erdman also relies upon the 2013 revalidation study of the Madison test, which said in general terms that firefighters' duties are "largely" the same, or "similar," across fire departments. These general observations would not have required the district court to find that the needs of the Madison fire department are a sufficiently close match to departments that use the IAFF test. The district court was not obliged to agree with Erdman on the IAFF test's local validity based on her affirmative evidence.

Even if Erdman had come forward with stronger affirmative evidence of the IAFF test's validity as applied to Madison, the city offered substantial evidence to the contrary. At trial, the city offered (1) "specific arguments in support of the two elements that plaintiff asserts are unnecessary as compared to the [IAFF test's] alternative elements—the ladder and pike pole," and (2) "evidence of numerous burdens associated with adopting the [IAFF test] as an alternative test."

The district court credited testimony regarding the time spent by the city's personnel and their expert consultants in developing those two elements of the 2014 Madison test that differed from the IAFF test. The city provided evidence that "those elements of the test were specifically designed to replicate the tasks Madison firefighters would be expected to execute in light of the equipment available to the Department at that time and of concerns about safety."

The district court's decisions about the effectiveness of and the intent behind those two events required it to weigh credibility and otherwise to weigh conflicting evidence. Such

findings call for great deference to the district court. *BRC Rubber & Plastics, Inc. v. Continental Carbon Co.*, 981 F.3d 618, 622 (7th Cir. 2020), quoting *Anderson v. City of Bessemer City*, 470 U.S. 564, 575 (1985) (“[W]hen a trial judge’s finding is based on his decision to credit the testimony of one of two or more witnesses ... that finding, if not internally inconsistent, can virtually never be clear error.”).

As to the ladder event, city witnesses explained why, in light of the equipment available to the department, a 20-foot ladder was chosen and why the lifting element challenged by the plaintiff had been removed after 2014.

For the pike pole test, the city also offered testimony that the requirement to stand 18 inches away from the ceiling being pulled down, rather than directly underneath it, reflects best practices for the safety of firefighters. Plaintiff did not show that these design differences were ineffective or merely pretextual. The district court’s finding that the IAFF test’s different version of these two events fails to serve the city’s legitimate needs is plausible in light of the entire record.

We addressed above the plausible basis in the record for the district court’s findings that “certain elements of the [Madison test] were designed specifically for Madison, in light of characteristics of the city, the Department’s equipment or other considerations, including safety.” 615 F. Supp. 3d at 900. Other evidence indicated that the Madison test is “a good predictor of outcome historically, as defined by a high passage rate out of the [fire training] academy.” *Id.* Chief Davis testified for the city that “very few people wash out because of physical ability.”

The evidence also permits an inference that the Madison test does a better job than the IAFF test at screening out applicants who are likely to wash out at later stages in the training process. Other fire departments have lower rates for hiring and retaining female firefighters (presumably using the IAFF test or others). Madison has a higher-than-average rate of hiring and retaining female firefighters. Those differences tend to support the district court's finding that Erdman failed to meet her burden at step three. After all, the ultimate concern here is not with how far women progress in the hiring process before being disqualified but with whether they can ultimately be hired and hold the jobs on a fair and non-discriminatory basis.

Evidence showed that "many, many" departments around the country use the IAFF test, which was developed in conjunction with ten leading fire departments in large cities across North America. Yet the city also offered evidence that the Madison fire department maintained a substantially higher rate of female firefighters than the national average; 14% in Madison in 2014 as compared to a national average of about 4%. The district court was careful to note that a "relatively strong record of hiring women more generally when compared to other fire departments around the country" did not excuse the Madison fire department from considering an alternative test. 615 F. Supp. 3d at 899. But even with this caveat, the district court was still persuaded that the most plausible inference was that Madison's high rate of female firefighters was traceable at least in part to the city's use of its physical ability test.

We are not as confident as the district court that Madison's lesser training wash-out rate and higher retention rate for

female candidates prove the Madison test is a better predictor of women's ability to train for and do the job successfully. Other possibilities may explain Madison's stronger record. These may include the fact that Madison's fire department was led by a woman for 16 years, that the city has a genuine commitment to hiring more women, that it may have a work environment that is welcoming to women, and provides role models, mentoring, and support for newly-hired women that may be missing in other cities. In a city the size of Madison, having a 365-person firefighting corps that is 10-14% women (36 to 51 women) rather than the 4% rate nationwide (14 or 15 women), may make the difference in whether a second or third woman is assigned to any given fire station or work shift, as opposed to one woman isolated by herself. Such differences matter to new hires in a historically male-dominated profession.

But the district court correctly described its logic on this point as one "possible inference" from the record. *Id.* at 900. Of course, on the city's motion for summary judgment, plaintiff Erdman as the non-movant was entitled to have the court draw any reasonable factual inferences in her favor. *Tolan v. Cotton*, 572 U.S. 650, 660 (2014). But not after the parties presented their evidence at trial. The district court was obliged to draw what it thought were the best inferences from the record. Again, we "must view the evidence in the entire record in the light which is most favorable to the finding." *Malachinski v. Comm'r of Internal Revenue*, 268 F.3d 497, 505 (7th Cir. 2001) (internal citations omitted). On this record, both parties are to one degree or another operating on certain unproven assumptions. We apply a deferential standard of review to mixed questions of law and fact, and recall that the burden at this third stage of the analysis was on plaintiff Erdman. We

are reviewing findings after a bench trial, and because the district court's analysis of the record on this issue is plausible, we must affirm. *BRC Rubber & Plastics*, 981 F.3d at 622.²

"[I]t is the trier of fact, rather than an appellate court, that resolves debatable factual issues.... It is enough to say that the district judge's handling of the disputes ... is thoughtful and reasonable." *Superl Sequoia Ltd. v. Carlson Co.*, 615 F.3d 831, 833 (7th Cir. 2010). Here, the district court's analysis of these issues was thoughtful and reasonable.

The judgment of the district court is AFFIRMED.

² We do not necessarily agree with every step of the district court's reasoning. For example, the district court counted the need for a transferability study as a burden relevant to the IAFF test's viability as an alternative hiring practice. We must not forget that at step three, we are considering whether a proposed alternative hiring practice would adequately serve the employer's legitimate needs. *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009), citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii) and (C). The purpose of a transferability study is to assess whether the job duties and equipment of Madison firefighters are sufficiently similar to those of the ten fire departments with whom the IAFF test was developed to ensure that the IAFF test would produce candidates qualified to be firefighters in Madison. But the substantial equivalence of the Madison and the IAFF test in producing minimally qualified candidates is itself one of the facts Erdman must prove to meet her third-step burden. Counting the need for and costs of a transferability study as weighing against plaintiff's proposed alternative hiring practice seems to create a Catch-22. But even apart from the transferability study question, ample evidence supports the district court's ultimate finding that Erdman failed to show the IAFF test would adequately serve the city's legitimate needs.