

UNITED STATES COURT OF APPEALS

FOR THE SECOND CIRCUIT

August Term, 2011

(Argued: November 15, 2011 Decided: July 30, 2012)

Docket Nos. 11-2184-cv, 10-2168-cv

M.O.C.H.A. SOCIETY, INC., MICHAEL BROWN, President, M.O.C.H.A. Society Inc.,
WILLIE BROADUS, ROBERT GRICE, ROBERT JONES, WALTER JONES, VICTOR MUHAMMAD,
WILLIAM RASPBERRY, JOHN TUCKER, individually and as representatives of African
American firefighters employed by the City of Buffalo and the Buffalo Fire Department,
who have been discriminated against by the promotional policies,

Plaintiffs-Appellants,

OTTO BREWER, individually and as representative of African American firefighters
previously employed by the City of Buffalo and the Buffalo Fire Department, who were
terminated pursuant to the Drug Testing Policy of the City of Buffalo Department of Fire,

Plaintiff,

v.

CITY OF BUFFALO, CITY OF BUFFALO, DEPARTMENT OF FIRE, CORNELIUS KING,
individually and as Commissioner, Department of Fire, JOHN D. SIXT, individually and as
Deputy Commissioner, Department of Fire,

*Defendants-Appellees.**

* The Clerk of Court shall amend the official caption as shown above.

M.O.C.H.A. SOCIETY OF BUFFALO, INC., EMANUEL C. COOPER, GREG PRATCHETT,
RUSSELL ROSS,

Plaintiffs-Appellants,

v.

THE CITY OF BUFFALO,

Defendant-Appellee.

Before:

KEARSE, WALKER, and RAGGI, *Circuit Judges.*

Appeals heard in tandem from judgments of the United States District Court for the Western District of New York (John T. Curtin, *Judge*) in favor of defendants on Title VII claims of race discrimination in the administration of the 1998 and 2002 promotional examinations for the position of fire lieutenant in the City of Buffalo Fire Department. See 42 U.S.C. § 2000e-2(a), (k). With respect to their disparate impact challenge to the 1998 examination, plaintiffs submit that the district court erred in finding, after a bench trial, that defendants had demonstrated that the examination in question was job related and consistent with business necessity based on a statewide job analysis in which Buffalo had minimally participated. With respect to their disparate treatment challenge to the 1998 examination, and their overall challenge to the 2002 examination, plaintiffs contend that the district court's award of summary judgment was based on an error of law insofar as it concluded that plaintiffs were barred from further litigating the job relatedness or business necessity of the two tests.

AFFIRMED.

JUDGE KEARSE dissents in a separate opinion.

THOMAS S. GILL, Esq., Frederick, Maryland, *for Plaintiffs-Appellants*.

JOSEPH S. BROWN (Joshua I. Feinstein, *on the brief*), Hodgson Russ LLP,
Buffalo, New York, *for Defendants-Appellees*.

REENA RAGGI, *Circuit Judge*:

Plaintiffs, M.O.C.H.A. (“Men of Color Helping All”) Society, Inc., and various named individuals, suing on behalf of themselves and as representatives of African American firefighters employed by the City of Buffalo (collectively, “M.O.C.H.A.”),¹ appeal two judgments entered by the United States District Court for the Western District of New York (John T. Curtin, *Judge*) on May 14, 2010, and May 24, 2011, in favor of the City of Buffalo, its fire department, and the fire department’s commissioner and deputy commissioner (collectively, “Buffalo”), on Title VII claims of race discrimination in making promotions to the rank of fire lieutenant based on 1998 and 2002 examinations that derived from a common statewide job analysis. Plaintiffs contend that, after conducting a bench trial, the district court erred in finding that Buffalo had successfully demonstrated that the test was job

¹ Although the individual plaintiffs are different in the appeals relating to the 1998 and 2002 examinations, we refer to all of plaintiffs collectively as “M.O.C.H.A.” for ease of reference. We refer to the entity M.O.C.H.A. Society of Buffalo, Inc., a fraternal organization of African American firefighters in Buffalo, as “M.O.C.H.A. Society” when discussing only the organization, not the plaintiffs as a whole.

related and consistent with business necessity, despite the disparate impact of the 1998 examination on African American applicants. See M.O.C.H.A. Soc’y, Inc. v. City of Buffalo (“M.O.C.H.A. I”), No. 98-cv-99C, 2009 WL 604898 (W.D.N.Y. Mar. 9, 2009). Plaintiffs submit that the district court further erred in awarding defendants summary judgment on the disparate treatment challenge to the 1998 test, see M.O.C.H.A. Soc’y, Inc. v. City of Buffalo (“M.O.C.H.A. II”), No. 98-cv-99C, 2010 WL 1875735 (W.D.N.Y. May 10, 2010), and on the overall Title VII challenge to the 2002 test, see M.O.C.H.A. Soc’y, Inc. v. City of Buffalo (“M.O.C.H.A. III”), No. 03-cv-580-JTC, 2010 WL 1930654 (W.D.N.Y. May 12, 2010), based on a determination that plaintiffs were barred from further challenging the job relatedness and business necessity of the two similarly derived promotional examinations.

A common question runs through these appeals, prompting us to hear them in tandem and now to decide them in a single opinion: Can an employer show that promotional examinations having a disparate impact on a protected class are job related and supported by business necessity when the job analysis that produced the test relied on data not specific to the employer at issue? We answer that question in the affirmative on the record developed in these related cases. While employer-specific data may make it easier for an employer to carry his burden at the second step of Title VII analysis, such evidence is not required as a matter of law to support a factual finding of job relatedness and business necessity. Where, as here, the district court hears extensive evidence as to how an independent state agency (1) determined, based on empirical, expert, and anecdotal evidence drawn from fire departments across New York and the nation, that the job of fire lieutenant, wherever

performed, involves common tasks requiring essentially the same skills, knowledge, abilities, and personal characteristics; and (2) developed a general test based on those findings, we conclude that the district court had sufficient evidence to make a preponderance finding that Buffalo’s use of that test to promote firefighters to the rank of fire lieutenant was job related and consistent with business necessity.² For this reason and others stated in this opinion, we affirm both challenged judgments in favor of defendants.

I. Background

A. The 1998 Examination

1. Test Development

In December 1997, Buffalo asked the New York State Civil Service Department (“Civil Service Department”), to create a promotional examination for the position of fire lieutenant in its fire department. See N.Y. Civ. Serv. Law § 23(2) (establishing that Civil Service Department shall prepare employment examinations for municipalities upon request). It was then standard practice for Buffalo to rely on the Civil Service Department for

² In reaching our conclusion that the district court did not clearly err in its job-relatedness finding, we note that Judge Curtin’s experience with discrimination within the Buffalo Fire Department spans almost thirty-five years. See, e.g., United States v. City of Buffalo, 457 F. Supp. 612 (W.D.N.Y. 1978) (holding that Buffalo Fire Department engaged in pattern or practice of hiring and promotion discrimination against African Americans, Latinos, and women), aff’d, 633 F.2d 643 (2d Cir. 1980); United States v. City of Buffalo, 721 F. Supp. 463 (W.D.N.Y. 1989) (modifying injunction imposed to remedy Buffalo Fire Department’s discriminatory hiring and promotion practices).

examinations for municipal civil service positions rather than to devise its own tests.³ In making the request, Buffalo provided the Civil Service Department with its fire department's most recent job specifications for the fire lieutenant position, the position's anticipated salary level, and promotion eligibility criteria.⁴

³ Buffalo no longer relies on the Civil Service Department for its promotional examinations, and instead uses private contractors through a bidding process.

⁴ The job specifications Buffalo provided for the job of fire lieutenant stated as follows:

DISTINGUISHING FEATURES OF THE CLASS

This is a first line supervisory position where incumbents are responsible for the activities of a fire company during an assigned shift. Responsibilities include directing the work of Firefighters at fires and in fire stations, evaluating their work performance and instructing them in new approved firefighting methods. Work is performed in accordance with established procedures and policies as outlined by the Fire Department. The class of Fire Lieutenant is distinguished from that of Fire Captain in that the latter is in charge when both he and the Fire Lieutenant are on duty. The Fire Lieutenant has complete charge of the activities of the fire company on all shifts and is in charge of operations at the scene of a fire in the absence of or pending arrival of a superior officer. All work is performed under general departmental regulations and incumbents directly supervise Firefighters under their command.

TYPICAL WORK ACTIVITIES

- Responds to all alarms assigned to his company while on duty;
- Directs the work of firefighters at scenes of fire and in station house;
- Assigns firefighters to lay out and connect hose lines and nozzles, direct hose streams, raise ladders and ventilate buildings;
- Inspects property at scene of fire to prevent re-ignition;
- Supervises the cleaning, checking and replacement of tools and equipment after a fire;
- Inspects personnel, station house, buildings, grounds and facilities to

Dr. Wendy Steinberg, an associate personnel examiner with the Civil Service Department, created the “Lower Level Fire Promotion” test series that was provided in response to Buffalo’s request. Steinberg testified that this test series was devised for use in promoting candidates to various firefighting positions, including fire lieutenant, in fire departments across New York. Indeed, municipalities across New York used the test series for that purpose.

To create the Lower Level Fire Promotion test series, Steinberg spent three years, from 1994 to 1997, conducting a job analysis of firefighters of all ranks from fire departments across New York. Based on this analysis, she designed examinations for each titled position. Steinberg’s job analysis had a dual focus: (1) the tasks firefighters perform and (2) the skills, knowledge, abilities, and personal characteristics (“SKAPs”) a person would be expected to possess on the very first day in a particular position. Steinberg testified as to how her job analysis was consistent with the joint standards of employment test design

-
- ensure conformity with departmental rules and regulations;
 - Examines fire trucks and equipment such as ladders and hose to ensure proper order and condition;
 - Inspects buildings and premises for fire hazards;
 - Personally supervises a wide variety of cleaning and maintenance tasks performed at the station;
 - Maintains discipline;
 - Makes periodic reports of personnel and activities;
 - Performs related duties as required.

J.A. 954, Dkt. No. 11-2184-cv. The salary was to be \$49,769, and applicants needed at least three years’ firefighting experience or one year’s experience as an assistant fire alarm dispatcher.

published by the American Psychological Association, the American Education Research Association, and the National Council on Measurement in Education, as well as with guidelines promulgated by the Equal Employment Opportunity Commission (“EEOC”).

In beginning her job analysis in 1994, Steinberg first collected job specifications from various New York fire departments for each submitted firefighting job title. From these specifications, she assembled a list of 190 identified tasks performed by firefighters of various ranks, which she reviewed with members of the Civil Service Department’s Fire Advisory Committee (“Fire Advisory Committee”), a panel of experts on the administration of fire departments. Steinberg then used the task list to create a statewide survey that, in April 1995, asked firefighters to rank each identified task according to how critical it was to the performance of firefighters’ specific jobs within their departments. With the Fire Advisory Committee’s assistance, Steinberg also created a second survey that, in October 1996, asked firefighters to rank listed SKAPs based on how critical they were to the respondents’ particular positions. Besides identifying the skills, knowledge, abilities, and personal characteristics necessary to perform the responsibilities of a specific job title, this survey was intended to provide a cross-reference for data obtained from the task survey.

The task and SKAP surveys were sent to every incumbent firefighter in New York, with the exception of those serving in New York City and Rochester.⁵ Steinberg followed

⁵ New York City’s and Rochester’s fire departments were not surveyed because they do not use examinations created by the Civil Service Department, but instead create their own tests internally. In the end, however, Steinberg did compare her Lower Level Fire Promotion test series plan against New York City’s and Rochester’s, as part of an effort to

up with non-responsive survey recipients in order to maximize the data obtained. In the end, of 5,934 task surveys sent out, 2,502 responses were received. Of those surveys, 2,994 were sent to firefighters serving in the state's twelve largest jurisdictions (other than New York City and Rochester),⁶ from whom 1,218 responses were received. Seven hundred ninety-five of the surveys were sent to firefighters holding fire lieutenant positions, with 316 responding. The responses in all three of these categories exceeded the numbers required by accepted statistical methodologies to establish 95% confidence in the survey results.⁷ Similarly, of 5,934 SKAP surveys sent, 1,604 responses were received, a number also sufficient to obtain 95% statistical confidence.⁸ By contrast, of 833 task surveys sent to Buffalo firefighters,

cross-validate the plan with fire lieutenant test plans from large jurisdictions across the country. See infra at 11 & n.9.

⁶ These twelve largest fire departments, in descending order of size, were Buffalo, Syracuse, Yonkers, Albany, Utica, White Plains, Troy, Binghamton, New Rochelle, Niagara Falls, Schenectady, and Mount Vernon.

⁷ Although Steinberg did not define 95% statistical confidence, we understand her to mean that there was a 95% probability that the survey results were not random, which makes it highly unlikely that they were the result of chance. See, e.g., Smith v. Xerox Corp., 196 F.3d 358, 366 (2d Cir. 1999), abrogation on other grounds recognized by Meacham v. Knolls Atomic Power Lab., 461 F.3d 134 (2d Cir. 2006); see generally Matrixx Initiatives, Inc. v. Siracusano, 131 S. Ct. 1309, 1319 n.6 (2011) (defining statistically significant results as those that are unlikely to be result of random error); Hans Zeisel & David Kaye, Prove It with Figures: Empirical Methods in Law and Litigation 85–88 (1997) (same).

⁸ Although Steinberg did not provide a source for her statistical confidence rates, M.O.C.H.A. has not challenged the accuracy or validity of her statistical calculations and conclusions, thereby waiving any such argument. See Norton v. Sam's Club, 145 F.3d 114, 117 (2d Cir. 1998) ("Issues not sufficiently argued in the briefs are considered waived and normally will not be addressed on appeal."). Rather, M.O.C.H.A. challenges only whether, as a matter of law, the district court could find that the statewide job analysis was suitable to the Buffalo Fire Department in the absence of other direct evidence.

only 68 responses were received, a number too low for the results to be reliable by themselves. Further, no Buffalo firefighter responded to the SKAP survey.

Upon receipt of survey responses from across New York, Steinberg grouped together the tasks identified as most critical to each firefighter title, including fire lieutenant. She performed a similar analysis of the SKAP survey responses and, with the assistance of other Civil Service Department staff, linked the most highly ranked SKAPs to corresponding highly ranked tasks. Steinberg then asked the Fire Advisory Committee to review and confirm the links drawn.

This process ultimately yielded six sets of task and SKAP categories that became the sub-test areas for the challenged fire lieutenant promotional examinations: (1) fire attack and suppression, (2) fire prevention, (3) rescue and first response, (4) understanding and interpreting written material, (5) training practices, and (6) supervision. These sub-test areas were approved by the Fire Advisory Committee.

Despite Buffalo's low response rate to the task and SKAP surveys, Steinberg determined that her statewide analysis was properly relied on in responding to Buffalo's request for a fire lieutenant promotional examination because of the overall consistency in the task and SKAP rankings of fire lieutenant respondents across New York. Despite the fact that responding fire lieutenants worked in different jurisdictions and in fire departments of varying sizes, Steinberg found a 90% correlation in the tasks they identified as critical to their job, in contrast to responses received from firefighters in other high-ranking positions, which showed more variance. Steinberg's conclusion that common tasks and SKAPs were

critical to the job of fire lieutenant across New York was buttressed by (1) the Fire Advisory Committee's review and approval of the tasks and SKAPs that she had identified for testing in a fire lieutenant examination, and (2) fire lieutenant promotional test plans from fourteen large fire departments across the United States, which were entirely consistent with the fire lieutenant test plan that Steinberg developed for general use by New York fire departments.⁹ Steinberg testified that obtaining such expert advice and cross-validating test plans against those of other jurisdictions are acceptable procedures under the joint standards of employment test design.

In addition, Steinberg invited subject matter experts from each of New York's fire departments to meet with her to discuss the questions to be included in the examination's sub-tests relating to fire lieutenants' firefighting tasks, i.e., the sub-tests assessing an applicants' fire attack and suppression, fire prevention, and rescue and first responder knowledge. The Buffalo Fire Department did not accept this invitation. Nevertheless, the multiple-choice questions that emerged from these discussions were then reviewed and approved by the Fire Advisory Committee. Multiple-choice questions for the remaining general sub-tests, i.e., understanding and interpreting written material, training practices, and supervision, were written by another Civil Service Department unit responsible for drafting

⁹ Those fourteen jurisdictions were: Los Angeles, California; San Jose, California; Denver City and County, Colorado; Dade County, Florida; Fort Lauderdale, Florida; St. Petersburg, Florida; Chicago, Illinois; Baltimore, Maryland; Prince George's County, Maryland; Reno, Nevada; New York, New York; Rochester, New York; and the District of Columbia.

general questions appearing on employment examinations across government agencies. Each sub-test carried the same weight in a candidate's final score, and Steinberg set the passing score at 66 correct answers out of 105 questions, which was lower than the maximum passing score of 73 under state law.

2. Test Administration

In providing Buffalo with its fire lieutenant promotional examination, the Civil Service Department also assumed responsibility for administering the test, which it did on March 14, 1998. The results showed a significant disparate impact. Of 179 white firefighters who took the test, 133 passed, a rate of 74.3%. Of 89 black firefighters who took the test, 38 passed, a rate of only 42.6%. Buffalo used these test results as its primary criterion in creating a fire lieutenant promotion list.

B. The 2002 Examination

Four years later, on April 6, 2002, the Civil Service Department again administered the Lower Level Fire Promotion test series for fire lieutenant applicants in the Buffalo Fire Department. Although new multiple-choice questions were written for the 2002 examination, the test was based on the same job analysis Steinberg performed for the Lower Level Fire Promotion test series developed in 1998, covered the same content as the 1998 examination, and was scored in the same manner. Plaintiffs allege and Buffalo admits that, in 2002, as in 1998, there was a significant disparity in the passing rates of white and black

applicants who took the examination.¹⁰

C. Procedural History of This Action

1. The Complaint Challenging the 1998 Test

M.O.C.H.A. Society, its president Michael Brown, and Buffalo firefighters Willie Broadus, Robert Grice, Robert Jones, Walter Jones, Victor Muhammad, William Raspberry, John Tucker, and Otto Brewer filed this action in February 1998, charging Buffalo with race discrimination both in its promotion policy and practice and in its enforcement of a drug-testing program. M.O.C.H.A. subsequently amended its pleadings to separate these claims into two complaints, with the October 2000 second amended complaint “B” here at issue alleging that Buffalo used a racially discriminatory examination in 1998 as the basis for promoting firefighters to the rank of fire lieutenant.¹¹ As M.O.C.H.A. alleged and later proved, African Americans passed the 1998 examination at a substantially lower rate than white candidates seeking promotion to fire lieutenant. Further, African Americans were generally under-represented in the Buffalo Fire Department’s upper ranks: while African Americans made up 30% of the city’s firefighters, they composed only 4% of fire officers at the rank of lieutenant or higher. Based on these facts, as well as allegations of intentional

¹⁰ Although the parties did not include the statistical data of disparate impact in the appeal record, it is stipulated that plaintiffs would have sustained their burden to prove the adverse disparate impact of the 2002 examination on African American applicants.

¹¹ Judgment entered on M.O.C.H.A.’s second amended complaint “A,” alleging race discrimination in Buffalo’s drug-testing policy, on May 31, 2012, and is not at issue on this appeal.

discrimination, M.O.C.H.A. charged Buffalo under Title VII with race discrimination on both disparate impact and disparate treatment theories. See 42 U.S.C. § 2000e-2(a), (k). As relief, M.O.C.H.A. sought an injunction voiding promotions made pursuant to the 1998 examination, as well as back pay, punitive damages, and attorneys' fees and costs.

2. The Complaint Challenging the 2002 Test

On July 30, 2003, M.O.C.H.A. Society and a different group of African American firefighters—Emanuel C. Cooper, Greg Pratchett, and Russell Ross—filed a similar complaint in the district court against the City of Buffalo, alleging that the 2002 examination was also discriminatory in violation of Title VII.

3. Bench Trial on the Disparate Impact of the 1998 Test

In 2008, the district court denied the parties' cross-motions for judgment as a matter of law and conducted a five-day bench trial to determine whether Buffalo was liable for the racially disparate impact of the 1998 examination. In its March 9, 2009 memorandum opinion, the district court ruled in favor of Buffalo, finding that, despite the disparate impact of the 1998 test on African American candidates, Buffalo had sustained its burden of proving that the test was job related and consistent with business necessity, whereas M.O.C.H.A. had failed to carry its rebuttal burden to show that an alternative promotional examination could have been used without disparate effect. See M.O.C.H.A. I, 2009 WL 604898, at *9–18.

4. Summary Judgment on the Remaining Claims

Buffalo subsequently moved for summary judgment on M.O.C.H.A.'s remaining claim of disparate treatment in connection with the 1998 examination as well as on the

complaint challenging the 2002 examination. Buffalo argued that the district court's trial finding of job relatedness and business necessity with respect to the 1998 examination precluded M.O.C.H.A. from re-litigating those issues with respect to either its claim of disparate treatment in 1998 or its challenge to the commonly derived examination administered in 2002. The district court agreed and concluded that, with the two tests' validity thus established, M.O.C.H.A. had failed to adduce sufficient other evidence to raise any triable issues of fact. Accordingly, on May 10, 2010, the district court granted Buffalo summary judgment on the disparate treatment claim pertaining to the 1998 test, and on May 12, 2010, it granted Buffalo summary judgment on plaintiffs' general challenge to the 2002 test. See M.O.C.H.A. II, 2010 WL 1875735, at *3–8; M.O.C.H.A. III, 2010 WL 1930654, at *2–6.

Upon the entry of final judgments and timely notices of appeal, this court heard the two appeals in tandem and now decides them together in this single opinion.

II. Discussion

In the first case we resolve in this decision, M.O.C.H.A. Society, Inc. v. City of Buffalo, dkt. no. 11-2184-cv, M.O.C.H.A. submits that (1) the trial evidence was insufficient to permit the district court to find that the 1998 examination on which Buffalo relied in making fire lieutenant promotions was job related and consistent with business necessity, thereby absolving Buffalo of liability for disparate impact discrimination; and (2) the district court erred in holding that plaintiffs could not re-litigate questions of job relatedness and business necessity to a jury in pursuing a disparate treatment challenge to promotions based

on the 1998 examination. In the second case that we resolve in this decision, M.O.C.H.A. Society of Buffalo, Inc. v. City of Buffalo, dkt. no. 10-2168-cv, M.O.C.H.A. contends that the district court erred in concluding that the adverse judgment regarding the 1998 examination collaterally estopped plaintiffs from pursuing their Title VII challenge to the 2002 examination.

We review the district court’s fact finding at trial for clear error and its legal conclusions de novo. See Gulino v. N.Y. State Educ. Dep’t, 460 F.3d 361, 381–82 (2d Cir. 2006). We review its awards of summary judgment de novo, resolving all ambiguities and drawing all permissible factual inferences in favor of the plaintiffs. See Burg v. Gosselin, 591 F.3d 95, 97 (2d Cir. 2010).

A. Docket No. 11-2184-cv: The 1998 Examination

1. Disparate Impact

Title VII makes it unlawful for an employer “to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual’s race, color, religion, sex, or national origin.” 42 U.S.C. § 2000e-2(a)(1). Promotion decisions are covered by this language. See, e.g., Ricci v. DeStefano, 557 U.S. 557, 578–85 (2011); Estate of Hamilton v. City of New York, 627 F.3d 50, 55 (2d Cir. 2010). To prove employment discrimination in violation of Title VII, a plaintiff need not show that an employer acted with the intent to discriminate. Rather, a prima facie violation may be established by statistical evidence showing that an employment practice has the effect of

denying members of a protected class equal access to employment opportunities. See New York City Trans. Auth. v. Beazer, 440 U.S. 568, 584 (1979); accord Gulino v. N.Y. State Educ. Dep’t, 460 F.3d at 382. Such a Title VII claim requires a plaintiff to make a statistical showing that a challenged employment practice has a disparate adverse impact on the protected class and, therefore, is usually referred to as a “disparate impact” claim. See 42 U.S.C. § 2000e-2(k); Gulino v. N.Y. State Educ. Dep’t, 460 F.3d at 382.

Here, there is no dispute that M.O.C.H.A. carried its prima facie burden to demonstrate disparate impact. Trial evidence showed that the passing rate for African Americans who took the 1998 examination (42.6%) was only 57.3% of the passing rate for whites who took the same test (74.3%). Under guidelines adopted by the EEOC, a selection rate for any Title VII protected class that is “less than four-fifths,” i.e., less than 80%, “of the rate for the group with the highest rate will generally be regarded . . . as evidence of adverse impact.” 29 C.F.R. § 1607.4(D). Consistent with our precedent, the district court properly deferred to these guidelines in finding M.O.C.H.A. to have carried its prima facie burden. See Teal v. Connecticut, 645 F.2d 133, 136–37 & n.6 (2d Cir. 1981) (stating that EEOC four-fifths rule merits great deference); see also Bushey v. N.Y. State Civil Serv. Comm’n, 733 F.2d 220, 225–26 (2d Cir. 1984) (holding that approximate 50% disparity between minority and non-minority candidates established prima facie case of disparate impact).

At that point, the burden shifted to Buffalo to show that the challenged 1998 test was “job related for the position in question and consistent with business necessity.” 42 U.S.C. § 2000e-2(k)(1)(A)(i); see Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975); Gulino

v. N.Y. State Educ. Dep't, 460 F.3d at 382. To carry this burden, Buffalo sought first to demonstrate that the 1998 examination measured content, rather than constructs. See generally Guardians Ass'n of N.Y.C. Police Dep't, Inc. v. Civil Serv. Comm'n (“Guardians I”), 630 F.2d 79, 92–93 (2d Cir. 1980) (distinguishing content validation, which determines whether employment examination’s testing of specific abilities is related to job, from construct validation, which determines whether employment examination’s testing of general mental processes and traits is related to job). Then, Buffalo sought to prove that the 1998 examination satisfied each of the factors that this court adopted in Guardians I for determining an employment test’s content validity. This required Buffalo 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. See id. at 95; accord Gulino v. N.Y. State Educ. Dep't, 460 F.3d at 384–85. The district court determined that the 1998 examination measured content, and that Buffalo satisfied each of the Guardians I factors, finding that the Civil Service Department’s statewide job analysis was suitable to creating a lieutenant examination for the Buffalo Fire Department, the examination writers were reasonably competent, the examination was related to and representative of the content of the Buffalo fire lieutenant position, and the scoring system was fair and selected those applicants most qualified to serve as fire lieutenant. See M.O.C.H.A. I, 2009 WL 604898, at *12–18.

This returned the burden to M.O.C.H.A. to show that a different test or selection mechanism would have served the employer's legitimate interests "without a similarly undesirable racial effect." Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 998 (1988) (internal quotation marks omitted); accord Gulino v. N.Y. State Educ. Dep't, 460 F.3d at 382. M.O.C.H.A. did not attempt to make such a showing. Rather, its trial strategy was limited to challenging Buffalo's ability to carry its burden at the second step of analysis.

That strategy having been unsuccessful in the district court, M.O.C.H.A. now argues on appeal that the district court clearly erred in its findings of job relatedness and business necessity. M.O.C.H.A. contends that, under Guardians I, Buffalo failed to show that the creators of the 1998 examination conducted a suitable job analysis. They further argue that the three generic sub-tests were not the product of reasonably competent test design and were not related to or representative of the content of the Buffalo Fire Department fire lieutenant position.

We review the district court's ultimate finding that the 1998 examination was job related, and its subsidiary findings regarding the individual Guardians I factors, for clear error. See Gulino v. N.Y. State Educ. Dep't, 460 F.3d at 386 (reviewing finding that employment test was properly validated for clear error); Guardians Ass'n of N.Y.C. Police Dep't, Inc. v. Civil Serv. Comm'n ("Guardians II"), 633 F.2d 232, 239, 241–42 (2d Cir. 1980) (reviewing district court's determination of examination's job relatedness for clear error), aff'd, 463 U.S. 582 (1983); see also Ass'n of Mexican-Am. Educators v. California, 231 F.3d 572, 584–85 (9th Cir. 2000) (en banc) (observing that all circuits to address

question have reviewed findings regarding test validation for clear error). Under that standard, we will reverse only where, ““on the entire evidence,” we are ““left with the definite and firm conviction that a mistake has been committed.”” Doe v. Menefee, 391 F.3d 147, 164 (2d Cir. 2004) (quoting Anderson v. Bessemer City, 470 U.S. 564, 573 (1985)). That is not this case.

a. Suitable Job Analysis

M.O.C.H.A. posits that the district court erred in finding that the 1998 examination was based on a suitable job analysis, i.e., an “assessment ‘of the important work behavior(s) required for successful performance and their relative importance.’” Guardians I, 630 F.2d at 95 (quoting 29 C.F.R. § 1607.14(C)(2)). Specifically, M.O.C.H.A. contends that because the challenged test was premised on the Civil Service Department’s statewide job analysis, in which Buffalo barely participated, the analysis amounted to only an “other cities guess” that the survey results were consistent with the job actually performed by lieutenants in the Buffalo Fire Department. Appellants’ Br. 35. M.O.C.H.A. charges that, as a matter of law, Buffalo was not entitled to rely on such guesswork and, by extension, that the district court clearly erred in similarly finding that the Civil Service Department’s job analysis was suitable to a promotional examination for the lieutenant position in the Buffalo Fire Department.

At the outset, we acknowledge that the Civil Service Department received minimal feedback from the Buffalo Fire Department in conducting its job analysis, and did not perform any on-site observations of that department or interview any of its members.

Further, at trial, Buffalo did not call any expert witness to opine that the results of the statewide job analysis correlated to the Buffalo fire lieutenant position. Nor did it present any direct evidence from a fact witness within the Buffalo Fire Department as to the responsibilities of a lieutenant in that department.

Buffalo's minimal participation in the Civil Service Department's three-year statewide job analysis of firefighter positions is perplexing. So too is Buffalo's strategic decision to defend against a disparate impact claim without calling either an expert or fact witness to link the lieutenant position within the Buffalo Fire Department to the Civil Service Department's job analysis of that position statewide. On such a record, it would have been within the fact finder's discretion to draw adverse inferences against Buffalo and to conclude that it had not carried its burden on the question of suitable job analysis. See Old Chief v. United States, 519 U.S. 172, 188–89 (1997) (recognizing that “triers of fact may penalize the party who disappoints them” in failing to introduce evidence to sustain his burden “by drawing a negative inference against that party” (internal quotation marks omitted)).

The issue before this court, however, is not whether a fact finder could have found against Buffalo on issues on which it carried the burden, but whether the fact finder here was required by law to do so. We conclude that he was not. In reaching that conclusion, we recognize that neither the Civil Service Department nor the district court may have been able to conclude as a matter of deduction that the statewide job analysis was suitable to the position of lieutenant in the Buffalo Fire Department without knowing more about that particular department. But that is not to say that the conclusion could not be reached as a

matter of induction. Application of the statewide job analysis to Buffalo was not a stab in the dark. Rather, it was based on a sound inference that, because reliable statistics showed that fire lieutenants across the state (and even the nation) shared the same critical tasks requiring the same critical skills, it was more likely than not that the same tasks and skills were critical to the fire lieutenant job in Buffalo. See Clark v. Astrue, 602 F.3d 140, 147 (2d Cir. 2010) (“In the civil context, a finding that X is more likely than not true is the equivalent to a finding that X is true.”). The law permits inferential fact finding even though it may be less certain than findings from direct evidence, see, e.g., Serricchio v. Wachovia Sec. LLC, 658 F.3d 169, 186–87 (2d Cir. 2011); Cifra v. Gen. Elec. Co., 252 F.3d 205, 217 (2d Cir. 2001), and even when the burden of proof is beyond a reasonable doubt, see, e.g., United States v. Abu-Jihaad, 630 F.3d 102, 135 (2d Cir. 2010), cert. denied, 131 S. Ct. 3062 (2011). Thus, despite Buffalo’s failure meaningfully to participate in the statewide analysis of the fire lieutenant position, we are satisfied that the district court still could find from the totality of the evidence that the Civil Service Department’s statewide job analysis was—more likely than not—suitable to identifying the tasks and skills relevant to the performance of that job in Buffalo.

Survey data warranting 95% statistical confidence showed that persons across New York with the title of “fire lieutenant” identified the same tasks as critical to their jobs regardless of the size or location of the fire department where they served. Indeed, 90% of surveyed New York fire lieutenants, when asked to rank specified tasks according to their criticality in the performance of the respondents’ jobs, provided virtually identical responses,

and those who departed did so only “slightly.” Trial Tr. 319. Such data made it highly likely that the job of fire lieutenant, wherever performed in New York, had the same critical tasks and required the same critical skills. Indeed, state survey data showed greater consistency across New York with respect to the position of fire lieutenant than with other high-ranking firefighter positions, where responses were more variable.

Thus, M.O.C.H.A. mischaracterizes the Civil Service Department’s job analysis when it contends on appeal that Steinberg simply relied on shared job titles to “guess” that the content of a job in one location was the same as the content of a job with the same title in another location. Rather, the trial record demonstrates that it was the high degree of common responses to task and SKAP surveys provided by persons holding the title of fire lieutenant that permitted Steinberg to conclude with a high degree of confidence that the analysis reliably applied to the position of fire lieutenant across New York.¹²

Further, the trial record shows that Steinberg tested the conclusion derived from the survey data by various means consistent with the joint standards of employment test design. Not only did she arrange for the test categories created from the survey data to be reviewed and approved by experts serving on the Civil Service Department’s Fire Advisory

¹² Insofar as Steinberg testified at trial that “I guess the evidence is that it would be highly unlikely that [Buffalo] would be different,” Trial Tr. 354 (emphasis added), one can discern even from a cold record that her “guess” was not a product of baseless speculation but, rather, a reasonable inference drawn from a job analysis showing that fire lieutenant tasks are essentially identical across the studied jurisdictions in New York.

Committee,¹³ she also compared the resulting six-part test plan with test plans from fourteen other large urban fire departments and found general consistency. Indeed, M.O.C.H.A.'s own expert testified that the joint standards permit both of these validation methods, although he disagreed that they were sufficient in this case to establish job relatedness. Further, as the district court noted, see M.O.C.H.A. I, 2009 WL 604898, at *14, Buffalo's own job specifications for the fire lieutenant position mirrored those of other New York departments that had participated in the Civil Service Department survey, for example, White Plains and Newburgh, jurisdictions with 170 and 61 firefighting positions, respectively.

From the totality of this evidence, we are satisfied that the district court could make a preponderance finding that the Civil Service Department's statewide job analysis for the position of fire lieutenant was suitable even to a jurisdiction such as Buffalo, which had participated only minimally in that analysis, and further, that the test developed from that analysis was job related to the lieutenant position as performed in the Buffalo Fire Department. In short, this is not a case where the record shows that one municipality simply relied on another's employment test without any evidence of a correlation between the two jurisdictions' circumstances. See generally EEOC v. Atlas Paper Box Co., 868 F.2d 1487, 1499 (6th Cir. 1989) (holding that employer failed to sustain burden by assuming jobs were

¹³ Buffalo's failure to adduce evidence that any committee member was familiar with the Buffalo Fire Department was properly considered by the district court in deciding what weight to give the Fire Advisory Committee's approval of Steinberg's job analysis for the position of fire lieutenant, a matter we have no reason to question on appeal. See Joseph v. N.Y.C. Bd. of Educ., 171 F.3d 87, 93 (2d Cir. 1999).

similar and that familiar “intelligence tests are always valid”); 29 C.F.R. § 1607.9(A) (prohibiting employer from using promotional examination based on “nonempirical or anecdotal accounts of selection practices or selection outcomes”). Here, substantial empirical evidence, reinforced by expert review and jurisdictional comparisons, showed that fire lieutenants across New York performed the same critical tasks and required the same critical skills, regardless of the location and size of their departments. While this may not absolutely foreclose the possibility that other tasks and skills might be relevant to the job of fire lieutenant in Buffalo, the law does not demand such a showing even when the standard of proof is beyond a reasonable doubt, see United States v. Reich, 479 F.3d 179, 190 (2d Cir. 2007), much less when it is a preponderance, i.e., when a party need only prove that its version of the facts is more probable than its adversary’s, see Clark v. Astrue, 602 F.3d at 147. Here, the record evidence was sufficient to permit the district court to find that it was highly unlikely that the fire lieutenant job in Buffalo required different tasks and skills from those identified in the statewide survey. This, in turn, permitted it to find that the statewide job analysis for fire lieutenants was suitable to Buffalo, notwithstanding its fire department’s minimal participation in the statewide survey.

M.O.C.H.A.’s arguments to the contrary are not persuasive. Insofar as M.O.C.H.A. faults the statewide job analysis because the Civil Service Department failed to secure sufficient survey responses from the Buffalo Fire Department, we reiterate our earlier conclusion that it was not clear error for the district court to find the job analysis suitable to Buffalo in light of (1) survey results convincingly showing that the job of fire lieutenant is

effectively the same across New York fire departments, (2) expert approval of the statewide fire lieutenant test plan, (3) the similarity between the statewide fire lieutenant test plan and test plans of large jurisdictions nationwide, and (4) the similarity between Buffalo's fire lieutenant job specifications and those of other New York fire departments whose members participated in greater number in the surveys.

Next, to the extent that M.O.C.H.A. suggests that in-person observations or interviews within the Buffalo Fire Department were necessary for a suitable job analysis, we think it misreads our precedent. In describing "a proper job analysis" as including "a thorough survey of the relative importance of the various skills involved in the job in question and the degree of competency required in regard to each skill," we have stated that such a survey "is conducted by interviewing workers, supervisors and administrators; consulting training manuals; and closely observing the actual performance of the job." Guardians II, 633 F.2d at 242 (internal quotation marks omitted). Although interviews and observations may be best practices, our endorsement did not impose an absolute requirement for every job analysis. In some circumstances, it may be possible to gather reliable job-specific information by other means, such as survey instruments sent directly to employees. Here, the Civil Service Department gathered data about the job of fire lieutenant by sending surveys to incumbent fire lieutenants statewide and receiving a sufficient number of responses to warrant 95% statistical confidence in the results. Further, these responses revealed statewide uniformity among responding fire lieutenants in identifying the tasks and SKAPs critical to their job. On these facts, and in the absence of any evidence to the contrary, we conclude that the survey methodology employed by Steinberg was adequate for a suitable job analysis.

Nor are we persuaded to fault the district court's finding because it was not informed by an expert opinion that the statewide job analysis was suitable to the Buffalo Fire Department. Such an expert opinion, like any direct evidence linking the Buffalo fire lieutenant's job to the statewide job analysis, would have strengthened Buffalo's defense. But expert opinion was not necessary as a matter of law. While the absence of expert testimony might carry more weight if Buffalo had been unable to rebut the testimony of plaintiffs' own expert, Dr. Kevin Murphy, that is not the case. Murphy asserted that Steinberg "simply assum[ed]" that the job of fire lieutenant is "pretty similar from one place to another without, to my knowledge, any detailed analysis, any analytic demonstration that that's true." Trial Tr. 263. But Steinberg herself refuted that charge by denying that she made any such assumption and, more important, by explaining that she relied on survey data showing a 90% correlation between fire lieutenants' tasks across jurisdictions. Because Steinberg provided a factual refutation to Murphy's criticism that a fact finder would be capable of comprehending without "special or peculiar training," we identify no clear error in the district court's decision to credit her testimony and to find her job analysis suitable to Buffalo even without an expert witness. Wills v. Amerada Hess Corp., 379 F.3d 32, 46 (2d Cir. 2004) (internal quotation marks omitted).

Similarly, we reject M.O.C.H.A.'s suggestion that Steinberg could not testify as to the challenged job analysis or 1998 examination without herself being qualified as an expert. As the district court observed, Steinberg testified to her personal knowledge regarding the statewide task and SKAP surveys, the results obtained therefrom, and the creation of the

1998 examination. Thus, M.O.C.H.A.’s argument is defeated by our decision in United States v. Rigas, 490 F.3d 208 (2d Cir. 2007), wherein we explained that “[a] witness’s specialized knowledge, or the fact that [s]he was chosen to carry out an investigation because of this knowledge, does not render [her] testimony expert as long as it was based on [her] investigation and reflected [her] investigatory findings and conclusions, and was not rooted exclusively in [her] expertise.” Id. at 224 (internal quotation marks and alteration omitted). In these circumstances, Steinberg’s testimony was admissible without regard to the limitations on expert opinions imposed by Fed. R. Evid. 702.

Nor do we identify clear error in the district court’s notation of the absence of evidence that Buffalo fire lieutenants perform different tasks or require different skills than do firefighters elsewhere in New York. See M.O.C.H.A. I, 2009 WL 604898, at *3 (describing Steinberg’s testimony). We do not understand the district court to have made this observation to shift the burden of proof from Buffalo or to allow Buffalo to carry that burden simply by reference to the lack of contrary evidence, either of which would have been legal error. See 42 U.S.C. § 2000e-2(k)(1)(A)(i); Albemarle Paper Co. v. Moody, 422 U.S. at 425. Rather, we understand this observation to reference the lack of any evidence impeaching Steinberg’s conclusion that, based on the survey data, the job of fire lieutenant involved the same tasks and required the same skills across New York, which would likely be true in Buffalo as well.

Finally, we do not identify clear error in the district court’s finding that the Civil Service Department’s job analysis was suitable despite Buffalo’s outlier size. The Buffalo

Fire Department is approximately twice the size of the Syracuse Fire Department, the next largest to participate in the statewide survey. Nevertheless, the district court could reasonably find that the job analysis was adequate even for a fire department as large as Buffalo's in light of the Civil Service Department's cross-validation of the test plan it created from its survey data with test plans from fourteen large fire departments nationwide. As Steinberg testified, the fourteen large departments' test plans "were all—at least 90 percent overlap—identical to the one that we gave to Buffalo and identical to each other." Trial Tr. 72.

In sum, we conclude that, despite Buffalo's failure meaningfully to participate in the Civil Service Department's statewide survey of firefighters, the district court was entitled to conclude that the ensuing analysis of the job of fire lieutenant was suitable even for Buffalo because survey results showed, with 95% statistical confidence, that fire lieutenants perform the same critical tasks and require the same skills, knowledge, abilities, and personal characteristics across New York. Indeed, the conclusion was reinforced by expert review and nationwide comparisons of the statewide test plan developed from the survey data. While the conclusion would have been stronger still if Buffalo had offered any expert or direct evidence linking its fire lieutenants' responsibilities to those identified in the statewide survey, we decline to hold that such evidence was required as a matter of law to permit the district court to make the requisite preponderance finding of a suitable job analysis.

b. Reasonable Competence and Content Relatedness

M.O.C.H.A. further contends that the district court could not find the challenged 1998 test job related to Buffalo's fire lieutenant position. In this respect, it maintains that the evidence was insufficient to show that the generic sub-tests in the 1998 examination—intended to assess (1) understanding and interpreting written material, (2) training practices, and (3) supervision—were the products of reasonably competent test design and were related to, and representative of, the fire lieutenant position. The record defeats these arguments.

(1) Reasonable Competence

The trial evidence was sufficient to establish that the generic sub-tests were the product of reasonably competent test design. In Guardians I, we stated that the reasonable competence of an employment examination's design can be called into doubt if (1) the examination was not created by professional test preparers, or (2) no sample study was performed to ensure that the questions were comprehensible and unambiguous. See 630 F.2d at 96. Here, the trial record shows that the generic sub-tests were written by professional test preparers. Steinberg testified that she relied on the Civil Service Department's units responsible for drafting cross-occupational test questions to write the generic sub-tests. Paul Kaiser, the Department's director of testing services who supervised the creation of the Lower Level Fire Promotion test series, detailed the process by which the Department drafts cross-occupational test questions and calibrates them according to specific job responsibilities and the promotion level at issue. Kaiser explained that, in doing so, the Department routinely consults with experts and people in the field at issue.

Buffalo offered no evidence that the Civil Service Department conducted a sample study to determine that the generic sub-test questions were comprehensible and unambiguous. But it did submit evidence that, in creating these sub-tests, the Civil Service Department employed cross-occupational questions from previous employment examinations, which had been screened for objections from past test administrators and takers. Whatever the advantages of a prospective sample study, we conclude that the Civil Service Department's consideration of such feedback to check the quality of cross-occupational questions, in combination with evidence that professional test preparers drafted them, was sufficient to support the district court's finding that the generic sub-tests were the product of reasonably competent test design.

(2) Content Relatedness and Representativeness

M.O.C.H.A. also contends that there was no proof that the generic sub-tests "measure[d] anything." Appellants' Br. 42. By this, we understand M.O.C.H.A. to be making three arguments, none of which is persuasive.

First, M.O.C.H.A. submits that Buffalo failed to introduce statistical evidence demonstrating that an applicant's success on the generic sub-tests is predictive of success as a fire lieutenant. This argument fails because the district court expressly found that the 1998 examination was content, not construct, validated, and that this content validation was an appropriate method for determining the examination's job relatedness. See M.O.C.H.A. I, 2009 WL 604898, at *12–13. M.O.C.H.A. does not challenge these findings on appeal.

In Guardians I, we explained that construct validation is “frequently impossible” because it requires “a demonstration from empirical data that the test successfully predicts job performance.” 630 F.2d at 92; see Gulino v. N.Y. State Educ. Dep’t, 460 F.3d at 384 (discussing greater difficulty of construct validation relative to content validation). By contrast, content validation does not require this predictive validation study, and only obligates a test-maker to show that the Guardians I factors were satisfied. See Guardians I, 630 F.2d at 95. Thus, even if a predictive validation study would have been preferable, see 29 C.F.R. § 1607.14(C)(5) (stating that, “[w]henver it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure” that has been content validated), we cannot conclude that its absence was fatal to Buffalo’s defense in light of the district court’s uncontested finding that content validation was appropriate for determining the 1998 examination’s job relatedness.

Second, M.O.C.H.A. asserts that the 1998 examination was not content related. See Guardians I, 630 F.2d at 97–98 (discussing content-relatedness requirement). In this sense, M.O.C.H.A. contends that the generic sub-tests were not reflective of a fire lieutenant’s tasks and, therefore, did not measure an applicant’s ability to serve in that position. The record shows, however, that in performing the job analysis, Steinberg conducted statewide surveys to determine the tasks a fire lieutenant performs and the SKAPs he would be expected to possess on the first day at that position. Based on the results of those surveys, Steinberg statistically ranked and linked the most critical tasks and SKAPs, which revealed that supervision, training, and understanding written material were related to the content of

the job. Cf. id. at 98 (describing process of identifying tasks and abilities in order to determine content relatedness). Further, based on those statistical results, consultation with the Fire Advisory Committee, and a comparison of nationwide test plans, Steinberg found that each sub-test area was important to the fire lieutenant position and should carry equal weight to ensure that the examination was content representative. Cf. id. at 99 (defining content-representativeness as proof that “the test measure[s] important aspects of the job, . . . but not that it measure all aspects, regardless of significance, in their exact proportions”). Having already determined that Steinberg’s underlying job analysis was suitable to the Buffalo fire lieutenant position, and in the absence of any evidence impugning Steinberg’s process of ranking and linking the task and SKAP surveys, we conclude that the evidence of Steinberg’s methodology was sufficient to support the district court’s findings that the 1998 examination was content related and representative.

Finally, to the extent M.O.C.H.A. posits that questions on the generic sub-tests were unrelated to the sub-test areas, i.e., that questions on the supervision sub-test had nothing to do with supervision, no record evidence demonstrates such disjunction. Once Buffalo showed that the Civil Service Department used reasonably competent means to construct the 1998 examination, the district court was permitted to infer that the generic sub-test questions were, in fact, related to the generic sub-test areas, which had independently been shown to be content related and representative. Further, although M.O.C.H.A. received the 1998 examination questions in discovery, it never contested the relationship between test questions and the subject matters they were supposed to cover. In the absence of any evidence

undermining the inference that, based on the Civil Service Department’s reasonably competent test design, the 1998 examination’s questions were related to the areas they were supposed to test, the district court did not err in finding the examination job related and consistent with business necessity.

Having identified no merit in M.O.C.H.A.’s various challenges to the finding of job relatedness, we conclude that the district court did not clearly err in determining that Buffalo had carried its burden at the second step of Title VII analysis. Accordingly, the district court properly entered judgment in favor of Buffalo on M.O.C.H.A.’s disparate impact challenge to the 1998 examination.¹⁴

¹⁴ Judge Kearse suggests that our affirmance of the district court’s ultimate job-relatedness finding will make it “virtually impossible” for municipalities to refuse to certify employment test results that have a proven disparate impact. See Dissenting Opinion, post at 5. Judge Kearse reasons that no municipality ever will have a “strong basis in evidence to believe it will be subject to disparate-impact liability,” the threshold necessary to discard test results and defend against a subsequent disparate treatment claim. Ricci v. DeStefano, 557 U.S. at 585.

Our court has only begun to define this “strong basis in evidence” standard. See United States v. Brennan, 650 F.3d 65, 110–14 (2d Cir. 2011) (setting forth factors for determining whether municipality established “strong basis in evidence” defense, including whether there is “objectively strong evidence of non-job-relatedness,” which can be demonstrated by less than preponderance of evidence); id. at 144–45 (Raggi, J., concurring in judgment) (expressing doubt that “strong basis in evidence” can be satisfied by less-than-preponderance showing). We have no occasion here to explore this issue further. We hold only that the district court, acting as fact finder after a bench trial, did not commit clear error in finding that a preponderance of the evidence showed that the 1998 examination was job related and consistent with business necessity. Whether such a relatively narrow and fact-dependent determination compels the broader legal conclusion that, at the time it certified the test results, the municipal employer lacked a “strong basis in evidence to believe it [was] subject to disparate-impact liability” is a question we leave for future courts to address.

2. Disparate Treatment

M.O.C.H.A. contends that the district court erred in awarding Buffalo summary judgment on plaintiffs' claim that promotions based on the 1998 examination reflected intentional disparate treatment of African American candidates for the job of fire lieutenant. See Robinson v. Metro-North Commuter R.R. Co., 267 F.3d 147, 160 (2d Cir. 2001) (noting that disparate treatment claim is distinguishable from disparate impact claim in that it depends on "the existence of discriminatory intent"). Specifically, M.O.C.H.A. argues that it should have been permitted to re-litigate before a jury the question resolved in favor of Buffalo at the bench trial, *i.e.*, whether the 1998 examination was job related. We are not persuaded.

By agreeing to a bench trial on the question of job relatedness, M.O.C.H.A. waived its right to re-try the issue before a jury. See generally Royal Am. Managers, Inc. v. IRC Holding Corp., 885 F.2d 1011, 1018 (2d Cir. 1989) (agreeing with majority of courts to have held that "participation in a bench trial without objection constitutes waiver of the jury trial right"). Insofar as M.O.C.H.A. contends that any waiver was limited to its disparate impact claim, it was nevertheless precluded by the law-of-the-case doctrine from re-litigating issues previously decided, which here included the district court's finding that the challenged 1998 examination was job related to the fire lieutenant position in Buffalo and consistent with business necessity. See Official Comm. of Unsecured Creditors of Color Tile, Inc. v. Coopers & Lybrand, LLP, 322 F.3d 147, 167 (2d Cir. 2003) (recognizing that district court discretion to revisit prior rulings is "subject to the caveat that where litigants have once

battled for the court's decision, they should neither be required, nor without good reason permitted, to battle for it again" (internal quotation marks omitted)).

Nevertheless, M.O.C.H.A. maintains that, even if the 1998 examination was job related, Buffalo was not entitled to summary judgment on a claim of disparate treatment because it used the test's results to make promotions even after it became apparent that the test had a disparate impact on African American candidates. This argument ignores that an employer cannot be held liable under Title VII, whether on a theory of disparate treatment or disparate impact, for conduct justified by business necessity. See Gulino v. N.Y. State Educ. Dep't, 460 F.3d at 383 (holding that employers can use employment tests having disparate impact, provided that tests are "demonstrably a reasonable measure of job performance") (quoting Albemarle Paper Co. v. Moody, 422 U.S. at 426)); see also United States v. Brennan, 650 F.3d 65, 93 (2d Cir. 2011) (holding that employer defending against disparate treatment claim has burden to articulate legitimate, non-discriminatory reason for its prima facie discriminatory practice).

In awarding summary judgment to Buffalo on M.O.C.H.A.'s disparate treatment claim, the district court relied on its previous finding that the 1998 examination was job related and, thus, justified by business necessity. M.O.C.H.A. failed to adduce any evidence indicating that Buffalo's non-discriminatory justification was a pretext for race discrimination. See M.O.C.H.A. II, 2010 WL 1875735, at *6-8; see also United States v. Brennan, 650 F.3d at 93 (discussing burden-shifting scheme for disparate treatment claims). Indeed, on appeal, M.O.C.H.A. points to no evidence that the district court overlooked or

misconstrued in this regard. Like the district court, we therefore conclude that summary judgment in favor of Buffalo was warranted because M.O.C.H.A. failed to adduce evidence giving rise to a triable issue of fact that “intentional discrimination was the defendant[s] ‘standard operating procedure’” in making promotions based on the challenged 1998 examination. Robinson v. Metro-North Commuter R.R. Co., 267 F.3d at 158 (quoting International Bhd. of Teamsters v. United States, 431 U.S. 324, 336 (1977)).

B. Docket No. 10-2168-cv: The 2002 Examination

In the second appeal that we decide today, M.O.C.H.A. argues that the district court erred in awarding summary judgment to Buffalo on plaintiffs’ Title VII challenge to the 2002 examination for fire lieutenant, based on the same statewide Civil Service Department job analysis underlying the challenged 1998 examination. The district court ruled that M.O.C.H.A. was collaterally estopped from challenging the 2002 examination by the district court’s rulings in favor of Buffalo on the 1998 examination. See M.O.C.H.A. III, 2010 WL 1930654, at *2–6.

“Collateral estoppel, or issue preclusion, prevents parties or their privies from relitigating in a subsequent action an issue of fact or law that was fully and fairly litigated in a prior proceeding.” Marvel Characters, Inc. v. Simon, 310 F.3d 280, 288 (2d Cir. 2002). The application of collateral estoppel to a given case is a question of law that we review de novo. See Faulkner v. Nat’l Geographic Enters. Inc., 409 F.3d 26, 34 (2d Cir. 2005). Here, it is not disputed that the validity of the 1998 examination was actually litigated and decided in the first case on appeal, and that the decision was necessary to the district court’s entry of

judgment in favor of Buffalo. Thus, we need only consider here whether (1) the issues in the two suits are identical, and (2) plaintiffs had a full and fair opportunity to litigate those issues. See Marvel Characters, Inc. v. Simon, 310 F.3d at 288–89.

M.O.C.H.A. contends that collateral estoppel is not warranted because the 1998 and 2002 examinations were not the same and, therefore, the issues with respect to the two tests are not identical. To be sure, the Civil Service Department changed the questions on the 2002 examination. But M.O.C.H.A.’s Title VII claims do not challenge the examinations’ questions. Rather, M.O.C.H.A. contests the underlying validity of the test in which those questions appeared, raising the same arguments that it put forward in its challenge to the 1998 examination. As a result, Buffalo’s defense to M.O.C.H.A.’s Title VII challenge to the 2002 examination would be exactly the same as its defense to M.O.C.H.A.’s challenge to the 1998 examination.¹⁵

M.O.C.H.A. perhaps could have raised a new challenge to the 2002 examination that was independent from the claims it pursued with respect to the 1998 examination. The district court found that M.O.C.H.A. waived any such challenge, however, stating that “the record reflects that the parties (and the court) have proceeded throughout the course of [the 2002 examination] litigation with the understanding that any determination made in

¹⁵ M.O.C.H.A.’s reliance on Suppan v. Dadonna, 203 F.3d 228 (3d Cir. 2000), is thus misplaced. Although the court there held that plaintiffs were not collaterally estopped from litigating a Title VII challenge to a promotion list different from the list at issue in an earlier lawsuit, the second lawsuit addressed theories of liability not present in the first suit. See id. at 232–33. Here, by contrast, plaintiffs are pursuing the exact same legal theories in their challenges to the 1998 and 2002 examinations.

M.O.C.H.A. I concerning the validity of the 1998 Lieutenant’s Exam would apply with equal force to the 2002 Exam.” M.O.C.H.A. III, 2010 WL 1930654, at *4. Indeed, in May 2007, M.O.C.H.A. proposed a stipulation to Buffalo that the “2002 Fire Lieutenant Examination . . . was prepared by the New York Department of Civil Service exactly as was the 1998 Fire Lieutenant Examination,” and that any decision regarding the validity of the 1998 examination would apply equally to the 2002 examination. Proposed Stipulation of Facts, M.O.C.H.A. Soc’y, Inc. v. City of Buffalo, No. 03-cv-580-JTC (W.D.N.Y. July 20, 2009), ECF No. 81, Attach. 1. M.O.C.H.A.’s conclusory assertions on appeal that it would have raised different issues with respect to the 2002 examination are contradicted by its prior litigation conduct and, in any event, are insufficient to sustain its burden on summary judgment. See Major League Baseball Props., Inc. v. Salvino, Inc., 542 F.3d 290, 310 (2d Cir. 2008).¹⁶

Nevertheless, M.O.C.H.A. posits that, by failing to respond to an interrogatory, Buffalo effectively admitted that the 2002 examination was invalid, distinguishing that examination from the 1998 version found valid by the district court.¹⁷ Because M.O.C.H.A.

¹⁶ To the extent that M.O.C.H.A. complains that the district court erred in denying M.O.C.H.A.’s belated request to depose Kaiser, we identify no abuse of discretion in light of plaintiffs’ failure to depose Kaiser during the lengthy pendency of the action. See Motor Vehicle Mfrs. Ass’n of U.S., Inc. v. N.Y. State Dep’t of Env’tl. Conservation, 79 F.3d 1298, 1308 (2d Cir. 1996) (reviewing denial of leave to conduct deposition only for abuse of discretion).

¹⁷ M.O.C.H.A.’s interrogatory read: “Do you contend that the 2002 Fire Lieutenant Examination was valid for selection of Fire Lieutenants as the term ‘valid’ was defined in Guardians Association of The New York City Police Department, Inc. v. Civil Service

failed to make this argument in the district court, we deem it waived on appeal. See Levitt v. Brooks, 669 F.3d 100, 104–05 (2d Cir. 2012). Even if the issue were properly before us, however, the record does not support M.O.C.H.A.’s tacit-admission contention. Buffalo did not completely fail to respond to the interrogatory; rather, it objected to its form. Further, the district court never granted M.O.C.H.A.’s motion to treat Buffalo’s purported silence as an admission. Finally, M.O.C.H.A.’s own actions in the district court do not signal its understanding of any admission by Buffalo. Had M.O.C.H.A. thought that Buffalo, by remaining silent, already had admitted that the 2002 examination independently violated Title VII, it would not have sought the requested stipulation that the issues in the 1998 and 2002 examination cases were identical. Accordingly, we conclude that M.O.C.H.A. presents identical issues with respect to both the 1998 and 2002 examinations.

M.O.C.H.A. separately maintains that collateral estoppel does not apply because the named plaintiffs in the suits challenging the 1998 and 2002 examinations are not identical and, therefore, plaintiffs in the latter action did not have a full and fair opportunity in the former action to litigate the validity of both tests. The argument is unconvincing where, as here, the named plaintiffs in the latter action had their interests “adequately represented” in the former action “by another vested with the authority of representation,” i.e., M.O.C.H.A. Society. Alpert’s Newspaper Delivery Inc. v. N.Y. Times Co., 876 F.2d 266, 270 (2d Cir.

Commission of the City of New York, 630 F.3d 79 (2d Cir. 1980)? If so, please explain.” J.A. 21. To this, Buffalo responded: “Defendant objects to this interrogatory on the grounds that it seeks a legal analysis rather than discoverable factual information and it seeks information protected by the attorney-client and work product privileges.” Id.

1989); accord Monahan v. N.Y.C. Dep't of Corr., 214 F.3d 275, 285 (2d Cir. 2000). As the district court found, “there can be no question that M.O.C.H.A. [Society] is the driving force behind the two lawsuits challenging the City [of Buffalo]’s use of the results of two successive administrations of the same Lieutenant’s Exam.” M.O.C.H.A. III, 2010 WL 1930654, at *5. Despite the absence of complete privity between the named plaintiffs in the two actions, “sufficient identity exists between the plaintiffs” for the district court to apply collateral estoppel and to bar litigation of the 2002 examination’s validity in light of its determination that the predecessor 1998 examination was valid. Alpert’s Newspaper Delivery Inc. v. N.Y. Times Co., 876 F.2d at 271.

Thus, because (1) M.O.C.H.A. seeks to re-litigate the identical issue with respect to the 2002 examination that was already decided against it in its lawsuit challenging the 1998 examination; (2) the plaintiffs in both lawsuits are sufficiently the same; and (3) the remaining collateral estoppel factors are not contested, plaintiffs were properly barred by collateral estoppel from challenging the validity of the 2002 examination.

III. Conclusion

To summarize, we conclude as follows:

1. On plaintiffs’ disparate impact challenge to the 1998 examination, the district court did not clearly err in finding that, despite the lack of direct evidence pertaining to the Buffalo Fire Department, Buffalo carried its burden to demonstrate the examination’s job relatedness by showing that the test derived from a valid statewide job analysis indicating that fire lieutenants across New York performed the same critical tasks and required the same critical skills.

2. On plaintiffs' disparate impact challenge to the 1998 examination, the district court did not clearly err in finding that the Civil Service Department exercised reasonable competence in designing the examination, and that the examination was both content related and representative.

3. On plaintiffs' disparate treatment challenge to the 1998 examination, the district court correctly concluded that plaintiffs could not re-litigate questions of job relatedness and business necessity decided against them at the bench trial of their disparate impact claim, and that M.O.C.H.A. had not established a genuine issue of material fact that Buffalo intentionally discriminated against African Americans by using the 1998 test results.

4. On plaintiffs' Title VII challenge to the 2002 examination, the district court correctly relied on collateral estoppel to grant summary judgment in favor of Buffalo because the only matters in dispute had been resolved against plaintiffs in the earlier challenge to the 1998 examination, and there was sufficient identity between the plaintiffs in both actions.

Accordingly, the judgments of the district court appealed in M.O.C.H.A. Society, Inc. v. City of Buffalo, dkt. no. 11-2184-cv, and M.O.C.H.A. Society of Buffalo, Inc. v. City of Buffalo, dkt. no. 10-2168-cv, are hereby AFFIRMED.

1 KEARSE, Circuit Judge, dissenting:

2 I respectfully dissent.

3 As can be seen from the Majority Opinion, the City of
4 Buffalo, for promotions to the position of fire lieutenant in its
5 Fire Department, used a 1998 test that had statistically significant
6 disparate impact on African Americans. Although Buffalo thus had
7 the burden of proving that the test was "job related for the
8 position in question and consistent with business necessity,"
9 42 U.S.C. § 2000e-2(k)(1)(A)(i) (emphasis added), it did virtually
10 nothing to carry that burden.

11 Buffalo simply requested a test and sent Dr. Steinberg a
12 sheet listing job specifications for the fire-lieutenant position.
13 Buffalo had not responded meaningfully to any of Dr. Steinberg's
14 requests for meaningful data in the preparation of the 1998 Exam.
15 It did not attend any of the exam-question formulation meetings to
16 which it was invited (see Trial Transcript ("Tr.") 86-87); and it
17 apparently did not encourage its incumbent fire lieutenants to
18 respond to the job-task survey circulated and to repeated follow-ups
19 by Dr. Steinberg (see, e.g., id. at 59, 72, 135). Dr. Steinberg's
20 overall project was to develop tests for all levels of fire
21 departments. The Buffalo Fire Department had 833 positions; Dr.
22 Steinberg's job-task survey elicited only 68 responses from Buffalo,
23 and she could not say that any one of them related to the lieutenant
24 position.

1 Dr. Steinberg also circulated a SKAP survey, i.e.,
2 questions about necessary Skills, Knowledges, Abilities, and
3 Personal characteristics. No one at any level of the Buffalo Fire
4 Department responded to this survey. (See Tr. 62, 66.) Dr.
5 Steinberg's 1995-1997 Fire Service Job Analysis report stated that
6 Buffalo "refused to participate at a meaningful level." Dr.
7 Steinberg testified that "Buffalo . . . wouldn't give [data] to me,
8 although I three times asked them to." (Tr. 72.)

9 Dr. Steinberg, who testified as a nonexpert, "presume[d]"
10 (Tr. 71) that data she received from Syracuse and Binghamton, whose
11 fire departments were far smaller than that of Buffalo, were also
12 applicable to Buffalo and that she had enough information to fashion
13 a test that would match the requirements of the fire lieutenant
14 position in Buffalo. But I have seen no evidence in the record from
15 which the trial court could verify her presumption. (Dr. Steinberg
16 also testified that she inferred that the needs of Buffalo would
17 match those of Albany (see id.); her report stated that Albany had
18 submitted no meaningful response.) The Majority refers to
19 "substantial empirical evidence," Majority Opinion, ante at 26; but
20 none of that evidence came from Buffalo. The Majority refers to
21 "jurisdictional comparisons" (id.); but several large cities in New
22 York State refused to participate in Dr. Steinberg's survey, and, in
23 any event, there were no possible comparisons with Buffalo. Dr.
24 Steinberg responded to questioning as follows:

25 Q. . . . [Y]ou couldn't look at how [the data from
26 other large jurisdictions] compared to the Buffalo data,
27 could you, because you didn't have any?

1 A. Nobody can look at how it compares to Buffalo
2 data because Buffalo didn't give us the data. That's why
3 we're talking about other large fire departments.

4 Q. So you're saying, as I understand it, I guess
5 Buffalo must be the same as the other large fire
6 departments?

7 A. I guess the evidence is that it would be highly
8 unlikely that they would be different. That was the
9 process I used throughout.

10 (Tr. 354 (emphasis added).)

11 After the test was prepared, it was sent to Buffalo for
12 approval. There is no evidence that anyone in or knowledgeable
13 about the Buffalo Fire Department even looked at it. The City's
14 Civil Service Director Olivia Licata testified that the City did
15 not, to the best of her knowledge, undertake any steps to validate
16 the Exam. (See Tr. 366-67.) Her records indicated that after she
17 (in her then-capacity as personnel specialist) sent the proposed
18 1998 Exam to the Fire Department for approval, she received no
19 response. (See id. at 361, 366.) Without a response, the apparent
20 routine was just to proceed with ordering the test. (See id. at
21 366.) And although Licata testified that she--a non-expert--would
22 usually check to see whether the exam got into areas that were not
23 in the job specifications (see id.), the City presented no evidence
24 that anyone more expert than Licata performed such an evaluation.

25 At trial, Buffalo presented no expert testimony to
26 validate the test. Nor did it present evidence that it had ever
27 hired an expert with respect to any phase of the test's conception
28 or preparation.

1 Thus, with no "expert opinion that the statewide job
2 analysis was suitable to the Buffalo Fire Department," Majority
3 Opinion, ante at 28, the district court allowed Buffalo to avoid
4 liability for its use of a racially discriminatory test on the
5 ground that Buffalo had proven content-validity

6 ■ without participating in the test preparation,

7 ■ without hiring an expert to advise it in advance whether
8 the test, prepared solely through the efforts of others,
9 would be suitably related to the job in question,

10 ■ without hiring an expert thereafter to evaluate the
11 content-validity of the test given and to testify to its
12 validity,

13 ■ without reference by the test's creator to any data to
14 substantiate her "guess" and her "presum[ption]" that the
15 data she received from others reflect Buffalo's
16 undisclosed needs,

17 ■ without presenting any evidence that any of Buffalo's
18 own knowledgeable personnel ever looked at the Exam
19 materials to determine whether the areas in which
20 questions were, or were to be, posed were material to the
21 job in question, and

22 ■ without making any attempt to show that the weighting
23 of the areas on the Exam reflected the requirements for
24 that position in Buffalo.

25 I am not persuaded that the evidence was sufficient to
26 support the district court's finding that Buffalo carried its burden
27 of proving that the test that was administered was job related for
28 the position of fire lieutenant in Buffalo.

29 And given the Supreme Court's recent decision in Ricci v.
30 DeStefano, 129 S. Ct. 2658 (2009), holding that a municipality may
31 not refuse to certify a test that has disparate racial impact unless
32 the municipality has a strong evidentiary basis for believing that

1 certification would subject it to disparate-impact liability, it
2 strikes me that this affirmance--allowing Buffalo to avoid liability
3 without having made any effort whatever to seek, verify, or defend
4 the test's validity--will make it virtually impossible for a
5 municipality not to certify for use a test that has clear
6 discriminatory impact.

7 Accordingly, I respectfully dissent.