

**NOT FOR PUBLICATION WITHOUT THE
APPROVAL OF THE APPELLATE DIVISION**

This opinion shall not "constitute precedent or be binding upon any court." Although it is posted on the internet, this opinion is binding only on the parties in the case and its use in other cases is limited. R. 1:36-3.

**SUPERIOR COURT OF NEW JERSEY
APPELLATE DIVISION
DOCKET NO. A-1777-23**

GABRIELE SPALLACCI, VICTOR
LORA, NOVAR VIDAL, LILLIAN
SANCHEZ, JUAN GARCIA,
PEDRO BORERRO, ROBERT
KLEIN, JUAN COSME, FELIPE
DIAZ, JOSE CASTELLANOS,
MARQUIS BROCK, MOHAMAD
DIABATE, ANGEL PARED,
VALERIA SANCHEZ-BERMUDEZ,
and ISABEL REYES,

Petitioners-Appellants,

v.

CIVIL SERVICE COMMISSION,

Defendant-Respondent.

Submitted May 28, 2025 – Decided August 22, 2025

Before Judges Sumners and Susswein.

On appeal from the New Jersey Civil Service
Commission, Docket No. 2024-916.

Law Offices of Steven A. Varano, PC, attorneys for
appellants (Albert J. Seibert, on the briefs).

Matthew J. Platkin, Attorney General, attorney for respondent (Donna Arons, Assistant Attorney General, of counsel; Craig S. Keiser, Deputy Attorney General, on the brief).

PER CURIAM

This appeal returns to us following our remand to the New Jersey Civil Service Commission to provide an explanation and interpretation of how the raw data it previously provided supported its decision to omit the last ten questions in scoring the February 23, 2019 promotional police sergeant exam because the questions created an adverse impact on racial minority examinees. See Spallacci v. Civil Service Commission, No. A-2369-20 (Aug. 7, 2023) (slip op. at 1-3). For the reasons that follow, we reverse the Commission's decision and invalidate the exam.

The parties are fully familiar with how we got to this point in their litigation, thus only a brief summary is necessary to provide context for our decision. In February 2019, petitioners—Gabriele Spallacci, Victor Lora, Novar Vidal, Lillian Sanchez, Juan Garcia, Pedro Borerro, Robert Klein, Juan Cosme, Felipe Diaz, Jose Castellanos, Marquis Brock, Mohamad Diabate, Angel Pared, Valeria Sanchez-Bermudez, and Isabel Reyes—took the police sergeant exam administered by the Commission. After the exam, the Commission's Division of Test Development, Analytics and Administration (TDAA) analyzed the

examination's raw data results and recommended that, in accordance with a consent decree reached with the United States Department of Justice (DOJ), as well as existing law, the last ten exam questions should not be scored because they disproportionately affected Black and Hispanic candidates, revealing a disparity in racial minority and non-racial minority candidates' performance. The Commission agreed and released the scoring results, excluding the last ten questions.

The exam's scoring was challenged by petitioners, thirteen of whom are racial minorities. After the Commission denied the challenge, petitioners appealed to us, "arguing the Commission's action was arbitrary and capricious, 'adversely impact[ing] the examinees that followed the instructions, managed their time properly, and completed the exam in the allotted time'" and, therefore the test results should be nullified. Id. at 3, 8.

In opposition, "[t]he Commission provided raw data consisting of several spreadsheets, outlining the 2019 exam and previous examination scores. These spreadsheets included, but were not limited to, mean scores for male candidates versus female candidates, as well as score breakdowns across different ethnicities." Id. at 6. We reversed and remanded, reasoning "the raw data supplied by the Commission to support its decision was indiscernible, lacking

explanation and interpretation regarding the adverse impact on racial minorities by scoring the last ten exam questions." Id. at 3.

We stressed:

The raw data affords neither petitioners nor us the ability to consider if scoring the final ten exam questions disparately impacted racial minorities, or whether, as petitioners suggest, the remedy adopted by the Commission unwittingly amplified rather than ameliorated the purported disparate impact it sought to correct. Under these circumstances, we cannot grant the Commission the deference we normally confer to an administrative agency. Accordingly, given the insufficient record before us, we do not pass judgment on whether the elimination of the ten questions was proper.

Remand is necessary for the Commission to provide an explanation and interpretation of how the raw data demonstrates the adverse impact on racial minorities by scoring the last ten exam questions.

[Id. at 10.]

On remand, the Commission cited the TDAA's October 4, 2023 letter, maintaining it explained how the raw data showed eliminating the last ten exam questions improved the test scores for minority candidates and reduced the adverse impact of those questions. In analyzing the exam results, the TDAA found that the last ten questions were omitted by examinees at much higher rates—eighteen to twenty-eight percent of the testing population omitted the last ten questions—when normally less than one percent of the testing population

omits a question. Omission rates for the last ten questions were even higher among Black and Hispanic examinees compared to White examinees, which increased the adverse impact on minority candidates. The removal of these questions, according to the Commission, "improved scores across the board," resulting in seventeen Hispanic and twelve Black examinees passing.

The TDAA, with the approval of the DOJ, utilized a statistical tool commonly referred to as "Cohen's d-value." The TDAA used the d-value¹ to measure the "effect size" between a test group—such as Black examinees—and a base group—such as non-minority examinees. It determined that a d-value of the last ten questions revealed a "moderate effect" between Black and White candidates, and also between Hispanic and White candidates. With the elimination of those questions, the TDAA determined: (1) Black examinees' d-values dropped from 0.914 to 0.847; and (2) Hispanic candidates' d-values decreased from 0.543 to 0.500. In terms of the entire exam, the TDAA found there was "at least a 'moderate effect' between the groups." The Commission reasoned it was well-founded as "there was a disparity for the [e]xamination as a whole . . . and that disparity was reduced by removing the last ten items."

¹ According to the TDAA, d-values measure the effect of a study, specifically, the difference between two groups' means in standard deviation terms. "Generally, a d-value of .2 is considered a 'small effect,' .5 is considered a 'moderate effect,' and .8 is considered a 'large effect.'"

The TDAA also explained that the last ten questions were likely not effectively measuring the intended knowledge, skills, or abilities (KSA), as many candidates may have been guessing. The Commission emphasizes the high rate of omitting the last ten questions suggests those questions were "ineffective." Thus, the TDAA concluded these questions were not serving their intended assessment purpose.

Petitioners argue, "[t]he random and arbitrary decision to remove the final ten questions unfairly punished those who followed the instructions and budgeted their time" and "rewarded those who spent additional time to respond to the more difficult questions preceding the final ten, irrespective of whether they even finished the examination." They argue the Commission undermined its own instructions and the exam, and thus significantly impacted their test performance. Petitioners contend the Commission improperly relied on the consent decree, which expired on November 22, 2014, to justify the omission of the questions on the January 16, 2016 exam.

Petitioners challenge the Commission's application of the four-fifths rule, measuring disparate impact. The Commission maintains Congress has set forth the four-fifths rule, 29 C.F.R. 1607.4(D), which provides that "[a] selection rate for any race . . . which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded by the [f]ederal enforcement agencies

as evidence of adverse impact." Petitioners point out the Commission did not discuss whether the prior three administrations of the police sergeant's exam violated the four/fifths rule because "[i]f the [] [r]ule was not violated in the prior administrations of the exam[], it calls into question whether the subject exam[] was simply an outlier and whether there is enough of a sample size to apply the [four/fifths] [r]ule." The inverse is also notable because "[i]f the [] [r]ule was applied to the prior administrations of the subject examination, and it was violated, the [Commission] did not set forth any explanation or information as to steps taken to mitigate the adverse impact and/or comply with the [] [r]ule."

Petitioners stress the Commission's assertion that "all candidates who took the exam were treated and scored equally," "is objectively false, as the elimination of the last ten questions adversely affected and impacted the scores of the candidates who followed the explicit and detailed instructions provided to the examinees and budgeted their time properly."

It is well settled that "courts will defer to an agency's grading of a civil-service examination except in the most exceptional of circumstances that disclose a clear abuse of discretion." Brady v. Dep't of Pers., 149 N.J. 244, 258 (1997). But here, we conclude the Commission is not entitled to that deference.

Both parties rely upon our decision in Rox v. Department of Civil Service, 141 N.J. Super. 465 (App. Div. 1976). In that case, there were sixty-one

candidates who were sitting for a promotional oral examination for police captain and were divided into groups and each group was graded by different personnel. Id. at 465-66. While all examinees were asked the same questions, their grades were calculated based on a subjective analysis of various characteristics measured by the different personnel, such as interpersonal relations and leadership qualities. Id. at 466. We concluded the administration of the exam was arbitrary, reasoning that because the examinees were divided into separate groups and scored by different personnel, using "a strongly subjective analysis," this was "violative of the spirit and purpose of the Civil Service rules." Id. at 467-68. We invalidated the oral exam results, citing N.J.A.C. 4:1-1.2, which requires the Commission to "assur[e] fair and impartial treatment for all applicants for employment and all employees in the classified service." Id. at 468. This standard was upended because "[t]he candidates were competing fairly with only those in their own particular group rather than with all candidates." Ibid.

We conclude Rox supports petitioners' appeal. The Commission's decision to eliminate the last ten questions because "between [eighteen percent] and [twenty-eight percent] [of the total testing population] omitted [the last ten questions]," is unpersuasive and arbitrary. This justification—standing alone—contradicts the exam's instructions, which emphasized:

The scoring of the written examination will be based on the number of correct responses. . . . [P]oints will not be deducted for wrong answers. Therefore, it is in the candidate's best interest to answer all questions. If the answer to a question is not known, choose the BEST choice. Candidates should budget their time so that they can respond to all questions within the allotted time.

[(Emphasis added).]

Moreover, the guidelines provided before the exam reiterated:

Prior to starting the exam, candidates will be informed as to the total number of items to answer and the total time allotted to complete the test. Candidates should budget their time so that they can respond to all questions within the allotted time.

[(Emphasis added).]

We agree with petitioners that the Commission's decision to omit the last ten questions after the test was taken undermines the agency's exam instructions because it essentially penalizes the examinees who allocated their time and provided answers to these questions. There is no indication the Commission explored alternatives to eliminating the last ten questions that did not punish examinees, such as petitioners, who diverted time away from the first seventy-five questions to ensure they completed the last ten questions. Petitioners were wrongfully penalized for following the instructions.

We are not persuaded by the Commission's decision that the last ten questions eliminated the Civil Rights Act of 1991's prohibition against the exam's discriminatory disparate impact. See Ricci v. DeStefano, 557 U.S. 557, 578 (2009). To evaluate such claims, an eighty percent standard from the Equal Employment Opportunity Commission as proof of disparate impact is applied. See 29 C.F.R. § 1607.4(D) (2008). Per the regulation, if the selection rate is less than eighty percent "of the rate for the group with the highest rate [it] will generally be regarded by the Federal enforcement agencies as evidence of adverse impact." Ibid. Crucially, the United States Supreme Court has held, "under Title VII, before an employer can engage in intentional discrimination for the asserted purpose of avoiding or remedying an unintentional disparate impact, the employer must have a strong basis in evidence to believe it will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory action." Ricci, 557 U.S. at 585 (emphasis added).

The Commission misapplied its discretion by eliminating the last ten questions because of their disparate impact on Black and Hispanic examinees. While the consent decree prohibits testing practices that adversely impact protected classes, the Commission's purported "remedy" of eliminating the last ten questions does not achieve that objective. As the data provided indicates and the Commission conceded, Black and Hispanic examinees were adversely

impacted by this exam. It is undisputed that after rescoring, Black examinees' d-values dropped from 0.914 to 0.847 and Hispanic examinees' d-values decreased from 0.543 to 0.500. Further, the Commission conceded, "there was a disparity for the [e]xamination as a whole . . . and that disparity was reduced by removing the last ten items." (Emphasis added). The Commission, however, did not remove the disparate impact, but rather "reduced" it. As such, its rescoring cannot be justified by "a strong basis in evidence to believe it will be subject to disparate-impact liability," Ricci, 557 U.S. at 585, as the new scores still had a "large effect" and "moderate effect" on Black and Hispanic examinees, respectively.

Moreover, petitioners correctly highlight that the Commission neglected to produce vital data to justify its rescoring. For instance, petitioners' right to appeal the exam's scoring procedure was stifled by the Commission's failure to disclose more complete data, detailing whether other specific questions adversely affected minority examinees. See N.J.S.A. 11A:4-1(e). Also, if other individual questions had a comparatively high d-value for Black and Hispanic examinees then the Commission would be more justified in omitting those particular questions. Invalidating questions that were substantively problematic, rather than invalidating questions based solely on their location in the exam, might remedy the disparate impact without penalizing those

examinees who followed the instructions and answered all questions. In effect, the Commission took the easy road, but not the fairest way to address the problem. Indeed, there was no evidence produced to verify whether its elimination was the most effective way to "reduce" the disparate impact on examinees. The Commission's failure to disclose such evidence is fatal as its decision to omit the final ten consecutive questions was arbitrary without further context.

We further conclude the Commission's justification for rescoring the exam because "the last ten items . . . could not be shown to have been effectively measuring the intended KSAs," is equally flawed. To support its remedy, the Commission notes that "minority groups were approaching the [twenty-five percent] threshold," indicating they were likely guessing and conclusively reasoning, "[i]f candidates are guessing on items, then the items are not properly assessing KSAs." This reasoning is unpersuasive.

To establish a disparate impact, the Commission must prove that these questions did not further a business necessity and were not related to job performance. See Ricci, 557 U.S. at 578; see also Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971). Petitioners correctly point out the Commission's justification lacks substance. For instance, the Commission did not detail how any of the eliminated questions failed to test an examinee's knowledge, skills,

or abilities. Indeed, the substance of these questions was not raised. The Commission's reasoning was premised on its assumption that when an examinee "guesses" an answer—because he or she is uncertain about the answer or is running out of time—that means there is something inherently wrong with the question itself. We conclude there is no logical basis for this reasoning and the Commission's decision to remove the final ten questions on this basis was arbitrary, capricious, and unreasonable. The Commission had the prerogative to present a shorter examination. But that decision should be made before the exam is administered, not after-the-fact, especially given the explicit instructions to answer all questions.

Because the integrity of the exam and its scoring has been undermined, we conclude that the exam results should be invalidated, and a new exam be administered. See Rox, 141 N.J. Super. at 468-69 (invalidating the oral portion of the examination that was deemed arbitrary).

To the extent we have not addressed any of the arguments raised on appeal, it is because we conclude they are without sufficient merit to warrant discussion. R. 2:11-3(e)(1)(E).

Reversed and remanded. We do not retain jurisdiction.

I hereby certify that the foregoing is
a true copy of the original on file in
my office.

M.C. Hanley

Clerk of the Appellate Division