

UNITED STATES DISTRICT COURT
DISTRICT OF MASSACHUSETTS

<hr/>)
BRUCE SMITH, PAUL JOSEPH, JOHN M.))
JOHNSON, ROBERT TINKER, MARTIN))
JOSEPH, KIM GADDY, BRIAN KEITH))
LATSON, LEIGHTON FACEY, MARWAN))
MOSS, and LATEISHA ADAMS,))
))
	Plaintiffs,)
))
v.)	CIVIL ACTION
)	NO. 12-10291-WGY
))
CITY OF BOSTON,))
))
	Defendant.)
))
<hr/>)

YOUNG, D.J.

July 26, 2017

MEMORANDUM & ORDER

I. INTRODUCTION

To understand why the Court here revisits and reconsiders rulings it made earlier in Smith v. City of Boston, 144 F. Supp. 3d 177 (D. Mass. 2015), it is necessary to understand the timing of my decision in Smith and how that decision may or may not conform to two other related yet distinct decisions -- Judge O'Toole's thorough opinion in Lopez v. City of Lawrence (Lopez I), No. 07-11693-GAO, 2014 U.S. Dist. LEXIS 124139 (D. Mass. Sept. 5, 2014) (O'Toole, J.), and its affirmance by the First Circuit, Lopez v. City of Lawrence (Lopez II), 823 F.3d 102 (1st

Cir. 2016), cert. denied, 137 S. Ct. 1088 (2017). The latter decision, of course, controls this Court's analysis.

All three decisions seek accurately to apply the law of disparate impact. At the most superficial level, the jurisprudence of disparate impact seeks fairly to ensure that employment decisions are made on genuine merit.

Recognizing that all employment tests are, by their very nature, discriminatory (after all, that's the whole purpose of testing -- to choose the few from the many), the plaintiffs must (first prong) prove that the test reveals a significantly disparate impact upon a lawfully protected minority -- significantly disparate impact because we don't want federal judges messing around with every employment test.

If the plaintiffs prove the first prong, the employer has the chance (second prong) to prove that the test vindicates itself through the business necessity of choosing on the basis of merit the best persons for the job.

Even if the employer prevails on the second prong, the plaintiffs get one last chance (third prong) -- to prove that there existed a test equal or better at identifying the best person for the job thus satisfying the employer's business necessity, which test was available to the employer and which test had a less disparate impact.

This is an elegant and nuanced matrix. The devil, of course, is in the details.

Lopez I, the first of these three related cases, commenced on September 11, 2007, with the filing of a complaint by a number of black and Hispanic patrolmen from various municipalities (including the City of Boston ("Boston")) challenging the civil service examination procedures for promotion to the rank of sergeant ("2008 sergeants' exam"). Drawn to Judge George O'Toole, this case came on for an eighteen-day bench trial commencing on July 12, 2010. When the trial concluded, Judge O'Toole took the case under advisement.

In February 2012, ten black police sergeants (the "Plaintiffs") in Boston commenced a substantially similar case before Judge Joseph Tauro. This case, the Smith case, challenged the police promotional exam from sergeant to lieutenant. When Judge Tauro took senior status, the case was transferred to this session on December 26, 2013.

In the meantime with Lopez I under advisement and Smith pending, Boston substantially revamped its police promotional testing procedures, adopting -- at significant expense -- many of the improvements for which both the Lopez I and Smith plaintiffs were contending.

On September 5, 2014, Judge O'Toole issued his full written opinion in Lopez I, finding that the 2008 sergeants' exam

imposed a significantly disparate impact on minority applicants, 2014 U.S. Dist. LEXIS 124139, at *48, and that the written portion of that exam could not alone support its validity "because it could not measure some skills and abilities (as distinguished from knowledge) essential to the position, such as leadership, decision making, interpersonal relations, and the like," id. at *60-61. Judge O'Toole went on to find that the Education and Experience portion of the examination saved it, albeit just barely. Id. The plaintiffs promptly appealed.

In Smith, the Plaintiffs alleged that the multiple-choice examination used by the Boston Police Department in 2008 to select and rank candidates for promotion from the rank of sergeant to lieutenant ("2008 lieutenants' exam") had a disparate impact on racial minorities and was invalid under Title VII of the Civil Rights Act of 1964. Smith, 144 F. Supp. 3d at 181. Boston responded that the exam did not have a disparate impact and, even if it did, was sufficiently job-related to be held valid. Id. at 180.

On December 15, 2014, at the outset of what proved to be a ten-day bench trial, the parties commendably moved into evidence the full trial record and exhibits from Lopez I. Then, for ten days, the Court heard lay and expert witnesses proffered by both sides, some of whom had not testified in Lopez I. See id. at 181. On November 26, 2015, this Court issued its opinion

concluding that the 2008 lieutenants' exam had a racially disparate impact and was insufficiently job-related to survive the Plaintiffs' challenge. Id. at 180-81. The Court thus imposed liability on Boston. Id. at 181.

Before engaging in extensive hearings concerning remedy, all parties sought time to explore settlement. After all, the challenged 2008 lieutenants' exam had long been out of use and the real nub of contention appeared to be the attorneys' fees due the Plaintiffs' counsel as prevailing parties.

Then, in a comprehensive opinion issued on May 18, 2016, the First Circuit affirmed Lopez I. Lopez II, 823 F.3d 102. As that court itself summarized: "[f]inding that the district court applied the correct rules of law and that its factual findings were not clearly erroneous, we affirm." Id. at 107.

Naturally, I read Lopez II with great interest. I was gratified to see that the First Circuit had unanimously concluded, as did Judge O'Toole -- and as had I with respect to the 2008 lieutenants' exam -- that the 2008 sergeants' exam had a significantly disparate impact on racial minorities. Id. at 111. On the sole issue where I had parted company with Judge O'Toole -- finding on different and additional evidence that business necessity could not justify use of the 2008 lieutenants' examination for the rank ordering of candidates for promotion -- the Court of Appeals had split 2-1 in reviewing

Judge O'Toole's findings as to the 2008 sergeants' exam. Id. at 122 (Torruella, J., concurring and dissenting). Most important, I detected no shift in the governing law in Lopez II from that I had applied to the facts I found in Smith. Nor would any shift be expected. Absent intervening Supreme Court precedent or legislative change, it is the practice in the First Circuit faithfully to adhere to the decisions of earlier panels of that court. See, e.g., Peralta v. Holder, 567 F.3d 31, 35 (1st Cir. 2009) ("We have held, time and again, that in a multi-panel circuit, prior panel decisions are binding upon newly constituted panels in the absence of supervening authority sufficient to warrant disregard of established precedent." (quoting Muskat v. United States, 554 F.3d 183, 189 (1st Cir. 2009))).

Whatever the potential legal effect of Lopez II on Smith, its practical effect was immediate. Settlement negotiations ceased. Now the parties sought an interlocutory appeal to settle once and for all the propriety of this Court's ruling on prong 2. This Court readily acceded to their wishes.

On October 11, 2016, the Court of Appeals gently but firmly rebuffed this gambit:

The district court issued its findings on liability in this case without the benefit of our subsequently-issued opinion in the case of Lopez v. City of Lawrence, 823 F.3d 102 (1st Cir. 2016). Since then, the district court has not yet purported to apply Lopez to the facts of this case.

For example, it has not stated whether and how its assessment of validity has taken into consideration the guidance we provided. Id. at 116-17. We therefore deny the petition without prejudice to renewal, if otherwise appropriate, after the district court has itself applied Lopez to this case.

J. United States Ct. Appeals 1, ECF No. 229.

In one sense, this order is both generous and courteous. It gives me first crack at applying Lopez II to my earlier legal analysis in Smith and making such analytic adjustments as may be necessary. Its tenor, however, suggests I may have missed something. Boston certainly thinks so.

In light of the First Circuit's order, this Court promptly held a status conference with the parties, Electronic Clerk's Notes, ECF No. 232, who subsequently briefed their positions on the effect of Lopez II on this Court's previous ruling in Smith, Pls.' Br. Ct. Appeal J., ECF No. 235; Pls.' Reply Br. Regarding Ct. Appeals J., ECF No 241; City of Boston's Br. Affect Lopez Ct.'s Liability Decision ("Def.'s Br."), ECF No. 236; City of Boston's Reply Br. Lopez's Affect Ct.'s Liability Decision ("Def.'s Reply"), ECF No. 242. Boston argues that Lopez II requires this Court to change its previous ruling by: (1) applying different legal standards to its prong 2 analysis, thus necessitating a different outcome, and (2) reaching prong 3 of the disparate impact inquiry.

II. THIS COURT'S PREVIOUS RULING

In these circumstances, this Court first conducts a brief, albeit rigorous and reflective review of what it has already done. In Smith, this Court examined the Plaintiffs' challenges to Boston's use of the 2008 lieutenants' exam to select and rank candidates for promotion from the rank of sergeant to lieutenant. 144 F. Supp. 3d at 180. The Court imposed liability on Boston, after concluding that the 2008 lieutenants' exam had a racially disparate impact and was insufficiently job-related to withstand the Court's disparate impact inquiry. Id. at 181.

In examining the evidence, this Court set forth the legal framework:

Under First Circuit case law, the plaintiff bears the burden of establishing a prima facie case of discrimination which consists of identification of an employment practice (in this case, the 2008 [lieutenants'] exam and promotions flowing therefrom), disparate impact, and causation.

If the Plaintiff meets this burden, the employer may either debunk the Plaintiff's prima facie case, or alternatively, may demonstrate that the challenged practice is "job-related and consistent with business necessity." If the employer demonstrates the latter, the ball bounces back into the plaintiff's court to demonstrate that "some other practice, without a similarly undesirable side effect, was available and would have served the defendant's legitimate interest equally well."

. . . .
. . . Under the first prong, the Plaintiffs must make a significant showing of actual disparate impact upon an identified protected minority

If the plaintiff can, however, make this showing, then under the second prong, the employer gets a chance to

demonstrate that the test in question is both job-related and consistent with business necessity. . . .

Even if the employer succeeds, however, the case is not over. Under the third prong, the plaintiff gets one more shot. If the plaintiff can demonstrate the availability of a testing program equally determinative of job performance, yet resulting in less disparate impact, the Court should fashion a remedy to secure the greatest degree of equal opportunity.

Id. at 182-83 (citations omitted). The Court then issued its findings of fact, id. at 185-91, thoroughly discussing the role of a Boston Police Department Lieutenant, id. at 185, pre-2005 job analyses and validation studies, id. at 185-88, the development and administration of the 2008 lieutenants' exam, id. at 188-89, the development and administration of the 2014 exam, id. at 189-90, and the results of the 2005 and 2008 lieutenants' exams, id. at 190-91. Next the Court set forth its conclusions of law. Id. at 191-211.

A. Prong 1

In its discussion of disparate impact (prong 1), id. at 191-200, the Court addressed the relevant data to consider in determining whether the Plaintiffs showed a significant disparate impact, id. at 192-94, whether or not to aggregate the data, id. at 194-95, and whether to use a one-tailed or a two-tailed test of statistical significance, id. at 195-98. The Court concluded that it would consider promotion rates, pass-fail rates, average scores, and delays in promotion to assess disparate impact, id. at 194, that it would not aggregate the

data, id. at 195, and that the Plaintiffs established disparate impact regardless of the Court's use of a one or two-tailed approach, id. at 198. The Court thus ruled that the Plaintiffs met their prong 1 burden to raise an inference of causation and demonstrate a prima facie case of disparate impact. Id. at 200.

B. Prong 2

The Court then assessed prong 2 of the disparate impact framework: job-relatedness and consistency with business necessity. Id. at 200-11. The Court noted that to prevail on prong 2, "[Boston] must convince the Court that the 2008 [lieutenants'] exam was both 'job related' for the position of Boston Police Department lieutenant and consistent with 'business necessity.'" Id. at 200 (quoting Jones v. City of Boston (Jones I), 752 F.3d 38, 53 (1st Cir. 2014) (quoting 42 U.S.C. § 2000e-2(k)(1)(A)(i))). The Court cited First Circuit precedent, stating:

to satisfy this second prong . . . "the [defendant] must show that its program aims to measure a characteristic that constitutes an 'important element of work behavior'" [and] "that the outcomes of [its challenged practice] are 'predictive or significantly correlated with' the characteristic described above."

Id. at 201 (quoting Jones I, 752 F.3d at 54 (quoting Albemarle Paper Co. v. Moody, 422 U.S. 405, 431 (1975))). Noting that the

Uniform Guidelines¹ "provide a sensible way of evaluating whether a given test . . . measures an important work characteristic, and whether the outcomes of that test are actually correlated with the characteristic measured," the Court "look[ed] to the Uniform Guidelines throughout its prong 2 analysis." Id.

Proceeding through its inquiry, this Court first held that the 2008 lieutenants' exam measured characteristics that are important elements of work behavior, id. at 201-02, because the job analyses on which the test was based "were sufficiently thorough and current so as to form solid ground on which to build a valid test." Id. at 202.

The Court then addressed if the exam results were predictive of or correlated with the important work behaviors. Id. at 202-10. The Court held that they were not, explaining that it:

ultimately agrees with [the Plaintiffs' expert] Dr. Wiesen that the evidence does not support the necessary inference that those who perform better on the exam will be better performers on the job, primarily because the exam did not test a sufficient range of [knowledge, skills, and abilities ("KSAs")], and there was no evidence that the exam was reliable enough to justify its use for rank ordering.

¹ 29 C.F.R. § 1607.4(C)(2). The Court explained that "Chapter 29 of the Code of Federal Regulations, section 1607, was published under the name of Uniform Guidelines on Employee Selection Procedures in 1978 by several government agencies to interpret how selection and testing and assessment should be conducted in accordance with the Civil Rights Act of 1964." Smith, 144 F. Supp. at 186 n.8 (citing 12/19/14 Tr. 23:6-18, ECF No. 166).

Id. at 203. Boston attempted to show job-relatedness through content validity, "an attempt to link the important KSAs of the job with the selection procedure." Id. The Court noted that the 2008 lieutenants' exam had two components: a multiple choice component, worth eighty percent of an examinee's score, and an Education and Experience score ("E&E"), worth twenty percent. Id. at 204. Relying on Dr. Wiesen's report, which correlated candidates' scores on the multiple choice section of the exam almost perfectly with their final score, the Court concluded that the E&E "had virtually 'no impact on the final exam scores,'" and thus excluded the E&E from the remainder of the Court's validity analysis. Id. (quoting 12/15/14 Tr. 58:19-21, ECF No. 151).

The Court then looked to test construction, id. at 204-06, and "f[ound] that the test construction process was inadequate to support the heightened validity requirement necessary to rank candidates," id. at 206. In reaching this conclusion, the Court noted that the test outlines showed that only two abilities appeared on the 2008 lieutenants' exam; Boston had sufficiently evaluated the test questions and their linkage to KSAs, but had "failed to conduct statistical analyses to ensure the quality of the test scores for the 2008 [lieutenants'] exam"; and that the record failed to address whether the test developer properly

recommended cut-off scores, rankings, bands, and weighting. Id. at 205.

Next, the Court addressed the 2008 lieutenants' exam's content. Id. at 206-08. Using the Uniform Guidelines' representative sample test, the Court found that the 2008 lieutenants' exam tested a sufficient range of the critical knowledge areas, but that the exam's near exclusion of any critical skills and abilities meant that "a high score on the 2008 [lieutenants'] exam simply was not a good indicator that a candidate would be a good lieutenant." Id. at 207-08.

Lastly, the Court discussed the use of the exam results to rank candidates. Id. at 208-10. The Court relied on the Guidelines' statements that "'evidence of both the validity and utility of a selection procedure should support the method the user chooses for operational use of the procedure, if that method of use has a greater adverse impact than another method of use,'" and that "'[e]vidence which may be sufficient to support the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the same procedure on a ranking basis.'" Id. at 208 (quoting 29 C.F.R. § 1607.5(G)). Noting that the Uniform Guidelines allow employers to validate rank-order exams with content validity, by establishing "'that a higher score on a content valid selection procedure is likely to result in better job performance,'" id.

(quoting 29 C.F.R. § 1607.14(C)(9)), the Court found that Boston failed to meet this standard because it neither tested a sufficient range of critical KSAs nor convinced the Court that the exam was valid, id. at 208-09. Accordingly, this Court concluded that Boston failed to show that a higher score on the 2008 lieutenants' exam would likely result in better job performance, id. at 210, and "h[eld] that even were the 2008 [lieutenants'] exam valid enough to be used as a screening tool, [Boston] failed to meet its burden of showing that the 2008 [lieutenants'] exam was sufficiently valid to be used as a basis for ranking candidates," id. at 211. Thus, the Court held that Boston did not meet its burden on prong 2 of the disparate impact inquiry, and the Plaintiffs won the case. Id.

III. ANALYSIS

Boston raises several challenges to this Court's prong 2 finding, essentially arguing that Lopez II effected changes in disparate impact law that mandate a reconsideration of this Court's previous decision. Boston also challenges this Court's failure to reach prong 3, arguing that the Court could not reject the exam without the Plaintiffs showing an equally valid, less discriminatory alternative. Id. at 12-18. The Court concludes that none of Boston's arguments have merit and upholds the original decision in Smith.

A. 2008 Lieutenants' Exam Validity (Prong 2 Challenges)

Boston argues that this Court ought reconsider its prong 2 analysis because Lopez II: (1) disposes of the Guidelines' representative sample inquiry in favor of a better than random selection test, (2) mandates that this Court include the E&E component in its assessment, and (3) holds that there is not a heightened validity requirement for Boston to use rank ordering. Def.'s Reply 5-11. These arguments are not convincing. Further, even were this Court to follow Boston's proposed test, Smith's conclusions would lead to the same result.

1. Reliance on the Guidelines

Boston argues that this Court used the wrong legal standard by relying on the Guidelines' representativeness inquiry, rather than Lopez II's purported "better than random selection" standard to determine the exam's validity. Def.'s Br. 6; Def.'s Reply 4.² The First Circuit certainly did not ban the use of the

² In advancing this argument, Boston comes close willfully to misreading Lopez II. Standing alone, the written multiple-choice lieutenants' examination is pretty clearly better than random selection yet the hard truth is that Boston's own experts recognize that such an examination disfavors minority applicants and Boston knows it. See Smith, 144 F. Supp. at 197 (citing 12/15/14 Tr. 113-14, ECF No. 161; 12/18/14 Tr. 45-46, ECF No. 165; Lopez I, 07/13/10 Tr. 82-85; Lopez I, 07/14/10 Tr. 43-48, 55, 59-60; Lopez I, 07/26/10 Tr. 30; Lopez I, 09/15/10 Tr. 58-59; Lopez I, 09/16/10 Tr. 110). Surely Boston is not here arguing that such an examination, standing alone, would pass muster. Every judge who has considered the issue has held to the contrary. Smith, 144 F. Supp. at 208, 210-11 (Young, J.); Lopez I, 2014 U.S. Dist. LEXIS 124139, at *60-61 (O'Toole, J.);

Uniform Guidelines' representative sample test. It would be surprising if it had since the Guidelines come from the Equal Employment Opportunity Commission and are due an appropriate degree of deference.

Still, Boston argues that this Court erred by applying the Guidelines as "binding legal standards" -- Boston asserts that Lopez II makes clear that the representative sample test cannot be used because the Guidelines do not have a quantitative measure for deciding whether or not a selection procedure tests a representative number of KSAs. Def.'s Br. 6-7.

This Court disagrees. Lopez II noted that although the Guidelines do not provide a quantitative measure to draw the line between representative and nonrepresentative samples of job performance, the Guidelines do "point to the qualitative understandings of these concepts generally accepted by professionals who evaluate [selection procedures]." 823 F.3d at 112. Accordingly, although there may not be a bright line to which reference can be made, expert testimony can still highlight on which side of a blurry line a selection device falls. In fact, Lopez II recognized that the testimony of Dr. James Outtz did just this -- "[Outtz] opined that the exams were

see also Lopez II, 823 F.3d at 113-15 (Lynch & Kayatta, JJ.); id. at 124-25 (Torruella, J., concurring in part and dissenting in part).

based on job analyses that validly identified the critical skills used by actual police sergeants and that the tests covered a 'representative sample' of the content of the job." Id. (quoting 29 C.F.R. § 1607.14(C)(4)). Here, in contrast, a different expert opinion on a different exam did not convince this Court that the 2008 lieutenants' exam measured a representative sample of relevant KSAs. Smith, 144 F. Supp. at 207-08.

Boston also argues that Lopez II declined to follow the Guidelines' technical requirements and instead established a lessened burden for the employer: a showing that the challenged exam is more job-related than random selection. Def.'s Reply 3. Boston posits that Lopez II "makes clear, 'in the absence of any quantitative measure of "representativeness" provided by the law,' the proper inquiry is not about representativeness, but whether the exam overall is more job-related than random selection would be." Def.'s Br. 6 (quoting Lopez II, 823 F.3d at 116). Indeed, Lopez II emphasized that "[t]he Guidelines quite understandably provide no quantitative measure for drawing the line between 'representative,' and nonrepresentative samples of job performance and behaviors." 823 F.3d at 112 (quoting 29 C.F.R. § 1607.5(B)). As Boston notes, the Lopez II court went on to reject the appellants' arguments, stating:

None of the remaining arguments advanced by the [appellants] seriously support any claim that the exams are not materially better predictors of success than would be achieved by the random selection of those officers to be promoted to sergeant. The parties' arguments, instead, focus on how much better the exams were. Do they test enough skills and knowledge? Do they weigh the answers in an appropriate, valid way? In finding Outtz persuasive on these points, the district court as factfinder did not clearly err.

Id. at 116-17.

Boston interprets the First Circuit's decision as a rejection of the argument that a test can be invalidated if it fails either to test a sufficient representative sample of skills and abilities or to meet a heightened standard of validity for rank ordering. Def.'s Reply 3-4. This, however, may well not be a fair interpretation -- in the quoted passage the First Circuit was merely deferring to the district court's role as the finder of fact on those points. Further, Lopez II upheld the district court's use of the representative sample test, noting:

The district court's opinion as a whole thus makes clear that the court trained its focus on critical and important knowledge, skills, and abilities called for by the job, and it did not clearly err by finding that a test that measured a large percentage of such critical and important KSAs was a test that was sufficiently "representative of important aspects of performance on the job." Our conclusion to this effect finds further support in the absence of any quantitative measure of "representativeness" provided by the law. Rather, the relevant aim of the law, when a disparate impact occurs, is to ensure that the practice causing the impact serves an important need of the employer, in which case it can be used unless there is another way to meet that need with lesser disparate impact.

We cannot see how it is an error of law to find that an exam that helps determine whether an applicant possesses a large number of critical and necessary attributes for a job serves an important need of the employer.

Lopez II, 823 F.3d at 115-16. Fairly read, this passage condones the district court's use of the representative sample test. It does not go so far as either to mandate or disapprove of that use as matter of law.

Nearly all of the circuits have utilized a representative sample test in examining content validity. See, e.g., Johnson v. City of Memphis, 770 F.3d 464, 478 (6th Cir. 2014), cert. denied sub nom. Johnson v. City of Memphis, Tenn., 136 S. Ct. 81 (2015); M.O.C.H.A. Soc'y, Inc. v. City of Buffalo, 689 F.3d 263, 281 (2d Cir. 2012); Equal Emp't Opportunity Comm'n v. Dial Corp., 469 F.3d 735, 742 (8th Cir. 2006); Allen v. City of Chicago, 351 F.3d 306, 309 n.3 (7th Cir. 2003); Zottola v. City of Oakland, 32 F. App'x 307, 311 (9th Cir. 2002); Nash v. Consolidated City of Jacksonville, Duval Cty., Fla., 837 F.2d 1534, 1539 (11th Cir. 1988), cert. granted, judgment vacated, 490 U.S. 1103 (1989), and opinion reinstated, 905 F.2d 355 (11th Cir. 1990). One imagines that, had the First Circuit been adopting a legal rule it would have acknowledged this body of law.

Boston argues that to satisfy its prong 2 burden, it need only establish that its test is a materially better predictor of

success than random selection. Def.'s Reply 6. This proposed standard, however, is not irreconcilable with the representative sample test. The aim of the representative sample test is to ensure that an exam tests for success in a specific job. Applying the representative sample test ensures that "the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated." 29 C.F.R. § 1607.5(B). In other words, a court ensures that a selection device evaluates characteristics important to job performance, rather than random attributes that may not correlate with success in that job. To be materially better than random at choosing applicants who will excel at a job, this Court can only imagine that the selection device would necessarily examine a large proportion of the KSAs needed to succeed at the position.

As discussed above, Lopez II did not reject the representative sample test as matter of law; and to assume that the First Circuit would change disparate impact law without so much as a comment seems somewhat at odds with reality. Accordingly, this Court declines to reconsider its use of the representative sample test from the Guidelines.

2. Rejection of the E&E Component

Boston argues that because Lopez II found that the E&E was useful for qualities important to a sergeant's daily

responsibilities, this Court should apply the E&E as part of its validity analysis. Def.'s Br. 9-10. In fact, this Court did consider the E&E in its ruling, but held the E&E did not rescue an otherwise invalid written exam. Smith, 144 F. Supp. 3d at 204.

The First Circuit discussed the E&E component of the 2008 sergeant's exam:

In Outtz's opinion, however, the addition of the E&E component effectively pushed the selection device as a whole across the finish line to show validity. It did this, according to Outtz, because the level and extent of work and educational experience and accomplishments listed by each applicant served as a useful, if imperfect, proxy for the kinds of qualities that were deemed to be important to a sergeant's daily responsibilities, yet were insufficiently tested by the examination's question and answer component. Outtz recognized that the gain in validity from the E&E component was, on its own, only marginal or "incremental." As the Officers stress, many of the attributes for which the E&E assigned points . . . were shared by all or most applicants. . . . when weighted to provide only 20% of the combined final score, it accounted for a range of only about 5% to 7% of a candidate's total score. Nevertheless, we cannot see how a rational factfinder could ignore the impact of the E&E, small or not, in evaluating the exam overall.

Lopez II, 823 F.3d at 113.

The evidence presented in Smith, however, varied from that available in Lopez I. In reaching its decision, this Court:

(1) relied on expert testimony that the E&E component failed to differentiate among candidates or demonstrate the KSAs necessary in a lieutenant, Smith, 144 F. Supp. 3d at 203-04; (2) had no evidence that incumbent lieutenants performed better on the

written exam, see generally id. at 207-10 (discussing the evidence presented to demonstrate exam validity); and (3) had no evidence to show that the E&E component was valid on its own, id. at 211 n.42.

These differences are crucial. In Lopez II, the E&E inched the exam over the line of validity due to its measuring "the kinds of qualities that were deemed to be important to a sergeant's daily responsibilities." 823 F.3d at 113. Here, however, the testimony does not establish that the E&E measured qualities important to a lieutenant's daily responsibilities. Further, even if the E&E assessed a lieutenant's important KSAs, Dr. Wiesen's testimony that the E&E "had virtually no impact on the final exam scores," Smith, 144 F. Supp. 3d at 204 (internal quotation marks omitted), persuades this Court that the E&E had so minimal an effect that it could not uphold the 2008 lieutenants' exam's validity. Accordingly, this Court declines further to address the E&E in its analysis.

3. Rank Ordering

Boston argues that this Court inappropriately applied a heightened validity requirement for rank ordering and that Lopez II holds that rank ordering furthers Boston's interest in eliminating patronage and intentional racism. Def.'s Br. 11. The First Circuit's statement is, however, dicta.

In Lopez II, the First Circuit stated:

Rank ordering furthers the City's interest in eliminating patronage and intentional racism under the guise of subjective selection criteria. Such a goal is itself a reasonable enough business need so as to provide some weight against a challenge that is unaccompanied by any showing that rank order selection itself caused any disparate impact in this case.

823 F.3d at 119. Boston asserts that this statement binds this Court, Def.'s Br. 11, and so Boston's business need to rank order supports its meeting the prong 2 burden in light of the Plaintiff's failure to set for evidence that rank ordering results in disparate impact, id. at 11-12. This Court, however, did find that rank ordering based on the 2008 lieutenants' exam had a disparate impact on minority applicants. Smith, 144 F. Supp. 3d at 199-200. In Smith, this Court stated, "the Plaintiffs have presented evidence of statistically significant disparate impact on . . . delay in promotions"; "if the eligibility list from the 2008 [lieutenants'] exam had not been extended due to the Lopez litigation, but instead had expired three years after its creation, as is typical, not a single black sergeant would have been promoted to lieutenant"; and "[a]ll of this evidence combined is enough for this Court to rule that the Plaintiffs have met their burden of raising an inference of causation and demonstrating a prima facie case of disparate impact." Id. Although these statements do not use the precise language "rank order selection caused disparate impact," this Court made it sufficiently clear -- through its

discussion of disparate impact in relation to delayed promotions -- that rank ordering resulted in disparate impact.

Boston goes on to argue that a prong 2 analysis does not depend on whether rank order selection increases disparate impact and that neither Title VII nor the Guidelines provide a quantitative requirement about how job-related a selection device has to be or how much better it need be for rank ordering. Def.'s Reply 9-10. "Employers should tailor a discriminatory hiring practice to a job-related risk, making sure to proportionally weigh the costs and benefits of accommodating that risk." Jake Elijah Struebing, Note, Reconsidering Disparate Impact Under Title VII: Business Necessity as Risk Management, 34 Yale L. & Pol'y Rev. 499, 507 (2016). Where a selection procedure not only has a disparate impact on a pass-fail basis, but also compounds that effect through use of rank ordering, each hiring decision carries an increased risk of a discriminatory result. Such heightened risk merits applying a more stringent validity requirement to ensure that the exam is sufficiently job-related to warrant the cost of potential discrimination. Accordingly, this Court did not err in applying a heightened validity requirement for rank ordering.

Further, even if Boston faced a lower bar to establish validity, it still failed to show that it met its burden. Although, as Boston argues, Def.'s Reply 10, this Court stated

that “[w]hat the Court can conclude from the 2008 [lieutenants’] exam is that those who excelled at the exam would exhibit superior levels of knowledge on the job, and that the 2008 [lieutenants’] exam differentiated among levels of candidates’ knowledge levels,” Smith, 144 F. Supp. 3d at 209, the Court also noted “that this is insufficient for predicting who will be a good police lieutenant,” id. In particular, this Court emphasized that testing only knowledge, rather than including other necessary skills and abilities, could not persuade this Court that those who did well would be better performers on the job. See id. Put another way, even were this Court to apply the “better than random” standard Boston advocates, Def.’s Reply 10, the Court concluded that it cannot presume that testing only knowledge will result in a better than random procedure for selecting candidates for promotion. As this Court stated, “the evidence does not support the necessary inference that those who perform better on the exam will be better performers on the job, primarily because the exam did not test a sufficient range of KSAs, and there was no evidence that the exam was reliable enough to justify its use for rank ordering.” Smith, 144 F. Supp. 3d at 203. Additionally, the Court later emphasized, “the Court cannot find that [Boston] met its burden on [adequacy of test construction]: too many skills and abilities were missing from the 2008 test outline,” id. at 205, and “[t]he Court

concludes that the 2008 [lieutenants'] exam did not sufficiently test for a representative sample of the critical KSAs," id. at 207. This goes to show that the Court held that even if Boston did not need to meet a heightened standard for rank ordering, it still failed to carry its burden to establish test validity.

B. Equally Valid, Less Discriminatory Alternative (Prong 3 Challenges)

Boston last argues that this Court cannot reject Boston's business justification unless there is some showing that there "exists an available alternative with less disparate impact that serves [Boston's] legitimate needs." Def.'s Br. 12. This Court, however, is confident in its understanding of the shifting burdens of the disparate impact framework: if the defendant fails to meet its burden of proof on prong 2, then the defendant loses, regardless of the plaintiffs' showing of an alternative.

Boston argues that Jones v. City of Boston (Jones II), 845 F.3d 28 (1st Cir. 2016), affirms Lopez II's lowered prong 2 standards and emphasizes the importance of prong 3. Def.'s Reply 1-2. This contention is hardly convincing. In Jones II, a number of police department employees brought a disparate impact challenge to the Boston Police Department's hair drug test, which they claimed was racially discriminatory. 845 F.3d at 30-31. Having held that the police department employees met

the first prong of the disparate impact inquiry in Jones I, the First Circuit examined the district court's grant of summary judgment on prongs 2 and 3, upholding the former and vacating the latter. See id. The First Circuit first turned to prong 2: whether the challenged test was job-related and consistent with business necessity. See id. at 32. Noting that "[t]he parties agree[d] that 'abstention from drug use is an important element of police behavior,' and is thus job related [and] that selecting police officers for retention or discharge based on that job-related behavior is consistent with business necessity," the court turned its analysis to whether the drug test was so unreliable that a reasonable juror could find that the test "did not meaningfully further the [Boston Police] Department's legitimate need for a drug-abstaining police force." Id. The court emphasized that the test had a high degree of accuracy, but that

there is no reason why a test need be anything near 100% reliable (few tests are) to be consistent with business necessity (keeping in mind that the presence of an alternative method that would have had less of a disparate impact will still be relevant under the third prong of the inquiry).

Id. at 33. The court then ruled that the Boston Police Department had clearly met its prong 2 burden to establish that the test was job-related and consistent with business necessity.

Id. at 33-34. Turning to prong 3 of the inquiry, the court

eventually held that police department employees could potentially succeed in showing that the Boston Police Department refused to adopt an alternative test with less of a disparate impact. See id. at 38. Accordingly, the circuit vacated the district court's grant of summary judgment on prong 3. Id.

Boston latches onto the Jones II court's focus on reliability, arguing that it evidences the First Circuit lessening the burden on employers in prong 2 while increasing the importance of prong 3. Def.'s Reply 2. Jones II, however, is distinguishable from the instant case on numerous grounds. First and foremost, in Jones II, both parties had essentially agreed that, if reliable, the drug test was job-related and consistent with business necessity. See 845 F.3d at 32. Here, in contrast, this Court specifically held that Boston failed to show that the 2008 lieutenants' exam was job-related and consistent with business necessity. Smith, 144 F. Supp. 3d at 211. Accordingly, an emphasis on reliability would be inappropriate. Second, Jones II addressed the validity of a test designed to look only at one thing -- drug abstention -- rather than a complex examination designed to test for the KSAs of promotion candidates. For this reason, the Jones II court's validity examination is distinguishable from this Court's business necessity analysis in Smith.

Further, Lopez II itself is consistent with the traditional burden-shifting framework of disparate impact. In Lopez II, the court summarizes the law of disparate impact, explicitly stating that a plaintiff who satisfies prong 1 will prevail either by the employer failing to meet their burden on prong 2, or by the plaintiff meeting their burden on prong 3. See Lopez II, 823 F.3d at 110-11. This outright statement of law warns against heeding Boston's call to collapse the inquiry of prongs 2 and 3.

IV. CONCLUSION

Both Lopez I and Smith are fact intensive cases. In Lopez I, Boston persuaded Judge O'Toole, albeit barely, that it ought prevail on prong 2. Smith is a different case, with a different evidentiary record, involving different expert testimony about a different and more demanding senior officer position. In Smith, Boston failed to convince me that it ought prevail on prong 2, an aspect of the case on which it bears the burden of proof. The Court of Appeals affirmed the district court's decision in Lopez I.

In view of this affirmance, Boston seems to be arguing that it is now legal error to reach a contrary result in a significantly different case.³ That's not how it works. Fact

³ Indeed, Boston apparently seeks to conjure up a hitherto unrecognized species of offensive issue preclusion, cf. Blonder-Tongue Labs., Inc. v. University of Ill. Found., 402 U.S. 313 (1971), based on the overlapping portions of the sergeants' and

finding is the province of the district courts. In Lopez II, the Court of Appeals did what it always does -- it carefully scrutinized the evidentiary record, giving due deference to the fact-finding role of the district judge, to see whether any of his conclusions were clearly erroneous. That's what it said it was doing. See Lopez II, 823 F.3d at 107-08. That's what it did. Should this Smith case be appealed, it will do the same for me.

Boston fails to offer any convincing argument as to why this Court ought disrupt its previous ruling in Smith. Accordingly, it declines to do so. The Court also takes this opportunity to note that Boston is not left without any useable test. Boston utilized an updated selection procedure in 2014. Def.'s Br. 4. Although Boston has represented to this Court that the 2014 exam resulted in a greater disparate impact and already faces legal challenges, id. at 4 & n.3, none of these issues are properly before this Court.

What is clear is that this case has gone on far too long. It is nearly a decade since the original Boston patrolmen brought suit in Lopez I. The tests that so engross us here are long out of date. In its order rejecting an interlocutory appeal, the First Circuit indicated it was amenable to

lieutenants' written exams and the fact that the same plaintiffs' counsel appears in both cases.

entertaining such an appeal once this Court had analyzed the effect of Lopez II on its earlier decision in Smith. It has now done so. Should the parties jointly move, within 14 days of the date of this memorandum, for an order authorizing application for a second interlocutory appeal, this Court will allow such motion. If not, this Court will promptly schedule hearings on remedy, settlement talks or no.

SO ORDERED.

/s/ William G. Young
WILLIAM G. YOUNG
DISTRICT JUDGE