

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----X

GULINO, ET AL.,

Plaintiffs,

96-CV-8414 (KMW)

OPINION & ORDER

-against-

THE BOARD OF EDUCATION OF THE
CITY SCHOOL DISTRICT OF THE CITY
OF NEW YORK,

Defendant.

-----X

WOOD, U.S.D.J.:

The question presently before the Court is a familiar one in this case: Does a teacher certification exam, developed by the New York State Education Department (the “SED”), discriminate against a class of African–American and Latino applicants for teaching positions in the New York City public school system, in violation of Title VII of the Civil Rights Act of 1964? This Court previously answered that question affirmatively regarding two different incarnations of the Liberal Arts and Sciences Test (the “LAST”), a certification exam no longer in use. *See Gulino v. Bd. of Educ. of City Sch. Dist. of N.Y.* (“*Gulino V*”), No. 96-CV-8414, ___ F. Supp. 3d ___, 2015 WL 3536694 (S.D.N.Y. June 5, 2015) (Wood, J.); *Gulino v. Bd. of Educ. of City Sch. Dist. of N.Y.* (“*Gulino III*”), 907 F. Supp. 2d 492 (S.D.N.Y. 2012) (Wood, J.). The Court must now answer the same question for the LAST’s successor: the Academic Literacy Skills Test (the “ALST”).

Plaintiffs contend that, like its predecessors, the ALST discriminates against the members of the class. The Court disagrees. Unlike the LAST, the ALST qualifies as a job related exam under Title VII. In 2010, in conjunction with its application for the United States Department of

Education’s Race to the Top program,¹ New York State adopted new federal and state pedagogical and curricular standards that redefined the role of teacher. The ALST was derived from those standards, and thus was appropriately designed to ensure that only those applicants who possess the necessary knowledge, skills, and abilities to teach successfully may be hired to do so in New York’s public schools. That conclusion relieves New York City of Title VII liability in this case.

I. NEW YORK STATE’S TEACHER LICENSURE EXAMINATIONS²

The SED requires the New York City Board of Education (the “BOE”) to hire only New York City public school teachers who have been certified to teach by the SED. *Gulino III*, 907 F. Supp. 2d at 498. The SED develops its certification requirements through a complex and largely internal process, which includes validation of tests to ensure that they do not have a discriminatory effect. *See generally* (ALST Tech. Manual [ECF No. 652]).

Beginning in 1993, the SED required teachers seeking certification to pass the first incarnation of the LAST (the “LAST-1”), a new test developed at the SED’s request by National Evaluation Systems (“NES”),³ a professional test development company. *Gulino III*, 907 F. Supp. 2d at 499–500. The LAST-1 “include[d] questions related to scientific, mathematical, and technological processes; historical and social scientific awareness; artistic expression and the

¹ New York’s application led the United States Department of Education to award New York State \$696,646,000, premised, in part, on the fact that New York “set forth a broad reform agenda, . . . including the development of new and more rigorous teacher preparation certification examinations.” (Wagner Decl. ¶ 18) (emphasis in original). These reforms included New York’s development of a comprehensive set of Teaching Standards, as well as New York’s adoption of federal Common Core Standards. NCS Pearson, Inc. (“Pearson”), which developed the ALST for the SED, relied heavily on these two standards documents in developing the ALST. *See infra* Part IV.A.

² For a more detailed discussion of the history of teacher licensure requirements in New York state, see *Gulino V*, No. 2015 WL 3536694, at *2, and *Gulino III*, 907 F. Supp. 2d at 498–500.

³ NES was acquired by Pearson in April 2006. *See Pearson Enters Teacher Certification Market by Acquiring National Evaluation Systems*, www.pearson.com (April 25, 2006), <https://www.pearson.com/news/announcements/2006/april/pearson-enters-teacher-certification-market-by-acquiring-national.html>. Pearson subsequently developed the exam at issue here—the ALST.

humanities; communication and research skills; and written analysis and expression.” (Foley Decl., Ex. I (“Clayton Decl.”) [ECF No. 377-3] at ¶ 4).

In 2004, the SED phased out the LAST-1 and introduced an updated version of the exam (the “LAST-2”). *See* (Dec. 8, 2009 Order [ECF No. 243] at 3). On May 1, 2014, the SED phased out the LAST-2 as well. *Gulino v. Bd. of Educ. of City Sch. Dist. of N.Y.* (“*Gulino IV*”), No. 96-CV-8414, 2015 WL 1636434, at *1 (S.D.N.Y. Apr. 13, 2015) (Wood, J.). In its place, the SED now requires prospective teachers to pass the ALST, an exam that purports to “measure[] a teacher candidate’s literacy skills . . . reflecting the minimum knowledge, skills, and abilities an educator needs to be competent in the classroom and positively contribute to student learning.” (Gullion Decl. [ECF No. 640] at ¶ 7). The SED contracted with Pearson to develop the exam. (*Id.* ¶ 6).

The ALST purports to measure a test taker’s “academic literacy” skills by assessing her knowledge, skills, and abilities (“KSAs”) within the domains of reading and writing. (*Id.* ¶¶ 7–8); (Wagner Decl. [ECF No. 638] at ¶ 38). The test has two components: a multiple-choice section, and an essay section. (Gullion Decl. ¶ 8). The multiple choice portion of the ALST contains five sets of eight questions, each set relating to a different reading passage. (ALST Tech. Manual at PRS012617–18). The reading passages are either literary (fictional) or informational (non-fictional). (*Id.* at PRS012617). Test takers must read each passage and answer questions that require careful analysis of the provided text. The essay portion of the ALST requires test takers to read two short reading passages and then construct several essays comparing and analyzing the passages. (*Id.* at PRS012617–18).

The SED requires prospective teachers to pass two exams in addition to the ALST: the Educating All Students test (the “EAS”), and the edTPA. (Wagner Decl. ¶¶ 32–38). According

to Pearson, “[t]he EAS measures skills and competencies that address: (i) diverse student populations; (ii) English language learners; (iii) students with disabilities and other special learning needs; (iv) teacher responsibilities; and (v) school-home relationships.” (*Id.* ¶ 36). The edTPA measures the performance of three pedagogical tasks: “(i) planning instruction and examination; (ii) instructing and engaging students in learning; and (iii) assessing student learning.” (*Id.* ¶ 35). Some teachers are also required to pass a Content Specialty Test (“CST”), (Gullion Decl. ¶ 6), an exam designed to “assess the specific knowledge and skills needed to teach specific subject matter in New York State public schools, such as mathematics, physics, chemistry, American Sign Language, Cantonese, Japanese, etc.” *Gulino V*, 2015 WL 3536694, at *2. Applicants must pass all required certification exams. *See* (Wagner Decl. ¶ 34).

II. PROCEDURAL HISTORY

The nineteen-year history of this case was recently set forth in *Gulino V*, as well as in the decisions in this case that preceded it.⁴ What follows is a condensed recounting of that history, as it relates to the current issues at bar.

A. The LAST-1 and LAST-2

Plaintiffs, who represent a class of African–American and Latino applicants for teaching positions in the New York City public school system, originally brought suit in 1996, three years after the LAST-1 was introduced. Plaintiffs alleged that the BOE had violated Title VII by

⁴ *See Gulino V*, 2015 WL 3536694 (finding the LAST-2 to be discriminatory under Title VII); *Gulino III*, 907 F. Supp. 2d 492 (finding the LAST-1 to be discriminatory under Title VII on remand); *Gulino v. N.Y. State Educ. Dep’t* (“*Gulino I*”), 460 F.3d 361 (2d Cir. 2006) (partially affirming and reversing Judge Motley’s original liability decision); *Gulino v. Bd. of Educ. of the City Sch. Dist. of N.Y.* (“*Gulino P*”), No. 96-CV-8414, 2003 WL 25764041 (S.D.N.Y. Sept. 4, 2003) (Motley, J.) (original liability opinion).

requiring applicants to pass the LAST-1,⁵ because the exam had a disparate impact on African-American and Latino test takers and did not qualify as job related.⁶

In 2012, this Court held that the LAST-1 had a disparate impact on the Plaintiffs and was not job related because it had not been properly validated by the State and NES. The Court thus concluded that the BOE had violated Title VII by hiring only teachers who were certified by the State (which certification required passing the LAST-1). *Gulino III*, 907 F. Supp. 2d at 516–23. Because the SED had retired the LAST-1 by the time the Court determined the test was discriminatory, the Court exercised its remedial authority to require that a “subsequent exam”—in this case the LAST-2—comply with Title VII. *See Gulino V*, 2015 WL 3536694, at *3 (citing *Guardians Ass’n of N.Y.C. Police Dep’t, Inc. v. Civil Serv. Comm’n of N.Y.* (“*Guardians*”), 630 F.2d 79, 109 (2d Cir. 1980)).

The Court appointed Dr. James Outtz to serve as a neutral expert to assess whether the LAST-2 had a disparate impact on African-American or Latino test takers—and if so, whether the exam qualified as job related. *See* (Apr. 29, 2014 Hr’g Tr. [ECF No. 428] at 55); (Oct. 29, 2013 Hr’g Tr. [ECF No. 403] at 4–8). Dr. Outtz concluded that the LAST-2 had a disparate impact on African-American and Latino test takers and did not qualify as job related, because it had not been validated properly. *See generally* (Outtz LAST-2 Report [ECF No. 549–1]). After

⁵ Plaintiffs initially sued the SED in addition to the BOE. *See generally Gulino I*, 2003 WL 25764041. However, the Second Circuit dismissed the SED from this case, holding that the SED is not an employer of public school teachers under Title VII. *Gulino II*, 460 F.3d at 370–79. For further discussion of this decision, see *infra* note 25.

Even though the SED is no longer a party to this suit, it was the SED, and not the BOE, which sought to defend the ALST as validly designed and implemented. Thus, this Opinion discusses the arguments put forward by the SED in depth, despite the fact that it is not currently a party to the suit.

⁶ In *Gulino II*, the Second Circuit held that the BOE could be found liable for New York State’s requirement because “Title VII preempts any state laws in conflict with it.” 460 F.3d at 380 (internal quotation marks omitted). The court stated that “even though BOE was merely following the mandates of state law,” by hiring only state-certified teachers, “it was nevertheless subject to Title VII liability.” *Id.* (internal quotation marks omitted).

a hearing on the matter, the Court agreed, and on June 5, 2015, held that the BOE had violated Title VII by hiring only teachers who had passed the LAST-2, among other tests.

B. The ALST

By the time the Court issued its decision concerning the LAST-2, however, the SED had retired the exam in favor of the ALST. *See Gulino IV*, 2015 WL 1636434, at *1 (concluding that the SED phased out the LAST-2 in favor of the ALST in 2014); *Gulino V*, 2015 WL 3536694, at *16 (finding the LAST-2 discriminatory on June 5, 2015). Accordingly, the Court again exercised its remedial authority to determine whether the ALST, as another subsequent exam, comported with the requirements of Title VII. *Gulino IV*, 2015 WL 1636434, at *3–5. As with the LAST-2, the Court appointed Dr. Outtz to assess the validity of the exam as a neutral expert. He concluded that the ALST had a disparate impact on African-American and Latino test takers, and that the exam did not qualify as job related because it had not been properly validated. *See generally* (Outtz ALST Report [ECF No. 645]). Beginning on June 22, 2015, the Court held a hearing at which witnesses on behalf of the SED and Pearson, as well as Dr. Outtz, testified about the validity of the ALST.

III. LEGAL STANDARD

Under Title VII, a plaintiff can make out a *prima facie* case of discrimination with respect to an employment exam by showing that the exam has a disparate impact on minority candidates. *See N.A.A.C.P., Inc. v. Town of E. Haven*, 70 F.3d 219, 225 (2d Cir. 1995). The defendant can rebut that *prima facie* showing by demonstrating that the exam is job related. *Id.* To do so, the defendant must prove that the exam has been validated properly. Validation requires showing, “by professionally acceptable methods, [that the exam is] ‘predictive of or significantly

correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.” *Gulino II*, 460 F.3d at 383 (quoting *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 431 (1975)).

In *Guardians*, the Second Circuit devised a five-part test for determining whether an employment exam, such as the ALST, has been properly validated and is thus job related for the purposes of Title VII:

- (1) “the test-makers must have conducted a suitable job analysis”;
- (2) the test-makers “must have used reasonable competence in constructing the test”;
- (3) “the content of the test must be related to the content of the job”;
- (4) “the content of the test must be representative of the content of the job”; and
- (5) there must be “a scoring system that usefully selects” those applicants “who can better perform the job.”

Guardians, 630 F.2d at 95.

Validation of an employment exam requires expertise that courts lack; it is “not primarily a legal subject.” *Guardians*, 630 F.2d at 89. Accordingly, to determine whether an employment exam was validated properly, a court “must take into account the expertise of test validation professionals.” *Gulino II*, 460 F.3d at 383. Courts also consider the Equal Employment Opportunity Commission’s Uniform Guidelines on Employee Selection Procedures (the “Guidelines”), 29 C.F.R. § 1607,⁷ which establish standards for properly validating an employment test. *Gulino III*, 907 F. Supp. 2d at 515. Although courts are not bound by the Guidelines, the Supreme Court has stated that the Guidelines are “entitled to great deference” because they represent “the administrative interpretation of [Title VII] by the enforcing agency.”

⁷ For clarity’s sake, the Court will denote sections of the Guidelines as “Guidelines § 1607.X,” rather than as “29 C.F.R. § 1607.X.”

Griggs v. Duke Power Co., 401 U.S. 424, 433–34 (1971). According to the Second Circuit, the Guidelines should be the “primary yardstick by which [to] measure [D]efendant[’s] attempt to validate [an employment exam]” because of their longstanding use in the field. *Gulino II*, 460 F.3d at 384.

IV. THE DEVELOPMENT OF THE ALST

Because courts must assess the validity of an examination based on the processes used and the decisions made to construct the test, what follows is a detailed description of the relevant events that preceded the development of the ALST, as well as the processes that contributed to its construction.

A. New York State’s Participation in “Race to the Top”

In July 2009, the United States Department of Education announced its “Race to the Top” initiative. The program offered federal funding to states that devised a comprehensive plan for educational reform that focused on four areas designated by the U.S. Department of Education: “[1] enhancing standards and assessments, [2] improving the collection and use of data, [3] increasing teacher effectiveness and achieving equity in teacher distribution, and [4] turning around struggling schools.” U.S. Dep’t of Educ., *Race to the Top Program Guidance and Frequently Asked Questions* 3 (2010), <http://www2.ed.gov/programs/racetothetop/faq.pdf>; see also generally U.S. Dep’t of Educ., *Setting the Pace: Expanding Opportunity for America’s Students Under Race to the Top* (2014),

https://www.whitehouse.gov/sites/default/files/docs/settingthepacertreport_3-2414_b.pdf.

New York submitted an application, which “set forth a broad reform agenda, . . . including the development of *new and more rigorous teacher preparation certification examinations.*” (Wagner Decl. ¶ 18) (emphasis in original). Based on that proposal, the U.S.

Department of Education announced in August 2010 that New York State would be awarded \$696,646,000 to implement its reform efforts. (*Id.* ¶ 19).

B. Developing the New York Teaching Standards

To begin that implementation process, New York’s Board of Regents directed the SED to develop what became known as the New York Teaching Standards (“Teaching Standards”). These standards “outline the requirements and expectations as to what teachers need to know before and after they become certified, and as they transition through their career from novice teachers, to experienced teachers, and eventually to master teachers.” (Wagner Decl. ¶ 21). According to the SED, the Teaching Standards “form the basis for assessing the minimum knowledge, skills and abilities required before a teacher enters a classroom.” (*Id.*)

The initial draft of the Teaching Standards was developed by the SED, and then revised by a working group, consisting of “teachers, principals, superintendents, faculty from teacher preparation institutions, as well as content area organizations, the National Board for Teachers, and parent-teacher groups.” (*Id.* ¶¶ 22–23). After an initial drafting, the Teaching Standards were “released to the field for comment through an on-line survey.” (*Id.* ¶ 23); *see also* (*id.* ¶¶ 24–27). The survey asked respondents to comment on both the clarity and appropriateness of the standards. (*Id.* ¶ 23).

Each Teaching Standard is defined at three levels of specificity. At the most general level, each Standard is formulated as a short statement. That statement is then broken down into “Elements,” which describe in more detail how teachers must meet the Standard at issue. (*Id.* ¶ 24, n.7). Each Element, in turn, is further fleshed out by “Performance Indicators” which describe “observable and measurable actions that illustrate each Element.” (*Id.*) For example, Teaching Standard I states: “Teachers acquire knowledge of each student, and demonstrate

knowledge of student development and learning to promote achievement for all students.” (ALST Tech. Manual, App. A [ECF No. 652-1] at PRS012705). One of the Elements of that Standard states: “Teachers demonstrate knowledge of and are responsive to diverse learning needs, strengths, interests, and experiences of all students.” (*Id.*) One of the Performance Indicators contained within that Element states: “Teachers vary and modify instruction to meet the diverse learning needs of each student.” (*Id.*)

C. Developing the ALST

The SED began its test development process by deciding that teachers would be required to pass at least three, and potentially four, certification exams before they could be licensed to teach in New York State: the ALST, the EAS, the edTPA, and in some circumstances, the CST. *See* (ALST Tech. Manual, App. M (“HumRRO Report”) [ECF No. 652-5] at PRS013019) (noting that the ALST was developed with the understanding that it would be one of several certification exams prospective teachers would need to pass, including the EAS, and the edTPA); (Wagner Decl. ¶ 28) (“As a result of the Race to the Top Application and after development of the [Teaching] Standards, the Board of Regents/NYSED asked their certification examination vendor, NES Pearson to develop, three new certification examinations for teacher candidates seeking their initial certificate - the ALST, the . . . EAS . . . and the edTPA . . .”). The SED made this decision before assessing the job tasks a New York State public school teacher performs day-to-day. (*Id.*)

Having decided that one of the required certification exams would be the ALST, the SED began the process of developing and validating the test.

There were two significant courses of action that contributed to the development of the ALST. The first, which the Court will term the “ALST Framework development,” involved the

creation of a framework that identified the KSAs the SED believed an incoming New York State public school teacher must possess to perform her job successfully. The ALST is intended to evaluate whether applicants possess these KSAs to a sufficient degree. The second process was the “job tasks analysis,” which identified and analyzed the job tasks and behaviors that New York State public school teachers perform in their daily duties. The SED contracted with Pearson to perform both of these processes. (Gullion Decl. ¶ 6). Pearson then sub-contracted with Human Resources Research Organization, (“HumRRO”), another test development organization, to perform the job tasks analysis. (Paullin Decl. [ECF No. 642] at ¶ 11).

i. The ALST Framework Development

Pearson initiated the ALST Framework development before HumRRO began its job tasks analysis. To start, “Pearson testing experts”⁸ studied two documents: the Teaching Standards, and the “New York State P–12 Common Core Learning Standards for English Language Arts and Literacy” (“Common Core Standards”). (ALST Technical Manual, at PRS012620). The Common Core Standards “define general, cross-disciplinary literacy expectations that must be met for students to be prepared to enter college and workforce training programs ready to succeed.” (ALST Technical Manual, App. L [ECF No. 652-4] at PRS012929). By adopting the Common Core Standards, New York specified, at least in part, the level of literacy that it expected New York State public school teachers to instill in their students. (Wagner Decl. ¶¶ 47–48). In essence, the Teaching Standards defined *how* New York’s teachers were expected to teach, while the Common Core Standards defined *what* they were expected teach.

⁸ Pearson describes these individuals as “a team of full-time senior test development staff that included literacy experts, former educators, psychometricians, and leaders with decades of experience developing assessment materials for educator credentialing in several states across the country.” (ALST Tech. Manual, at PRS012621).

Pearson’s testing experts used these two documents to develop an initial draft of the ALST Framework. This framework identified two KSAs⁹ that Pearson’s testing experts believed a teacher needed to successfully perform her job: “Reading” and “Writing to Sources.” *See* (Gullion Decl. ¶ 9). Pearson describes these KSAs as “specific areas of knowledge and skills which New York State job incumbents¹⁰ have determined are important to the critical tasks performed by teachers in the classroom, and thus have been determined to be the areas on which candidates are assessed by the ALST.” (*Id.*) (footnote added).

The Framework included “Performance Indicators,” which “provide further details about the nature and range of the knowledge and skills covered by the competencies.” (*Id.*) These indicators were “intended to suggest the type of knowledge and skills that may be assessed by the questions associated with each competency.” (*Id.*) For example, some of the Performance Indicators for the KSA of “Reading” are: “determines what a text says explicitly,” and “analyzes the development of central ideas or themes of a text.” (ALST Tech. Manual, App. I [ECF No. 652-2] at PRS012914–15). “Writing to Sources” Performance Indicators include: “evaluates the validity of reasoning used to support arguments and specific claims in a text,” and “anticipates and addresses a possible counterclaim.” (*Id.* at PRS012915–16).

Pearson’s draft Framework was reviewed by the SED, and then provided to two committees of educators for review. The Bias Review Committee (the “BRC”) was charged with “ensur[ing] that all test materials [were] free from bias and [were] fair and equitable for all

⁹ The SED, in its materials, often refers to these KSAs as “competencies,” but also sometimes describes them as KSAOs, which stands for knowledge, skills, abilities, and “other characteristics.” *See, e.g.*, (Gullion Decl. ¶¶ 8–9) (using the term “competencies” instead of “KSAs”); (Paullin Decl. ¶ 15) (defining KSAOs). To be consistent with the Court’s *Gulino V* Opinion, the Court will use the term “KSA” in place of either “competency” or “KSAO.”

¹⁰ Of course, at this point in the framework development, job incumbents had not yet made these determinations; the framework was initially developed by “Pearson testing experts,” not job incumbents. *See* (ALST Tech. Manual, at PRS012621).

candidates.” (ALST Tech. Manual, at PRS012622). The goal of the BRC was to “exclud[e] language or content that might disadvantage or offend an examinee because of her or his gender, race, religion, age, sexual orientation, disability, or cultural, economic, or geographic background,” and “includ[e] content that reflects the diversity of New York State.” (*Id.* at PRS012622). The BRC was made up of practicing educators with a special emphasis placed on “recruiting BRC members representative of the regional, gender, and ethnic diversity of New York State.” (*Id.* at PRS012623). Although the SED states that the BRC consisted of twenty-four educators in total, only four of them actually participated in the review of the ALST Framework. (*Id.* at PRS012623–24); (ALST Tech. Manual, App. K [ECF No. 652-3] at PRS012920–PRS012921). Of those four BRC committee members, two identified as African-American, one identified as Latino, and one identified as Asian-American. (Pearson Ex. 60 [ECF No. 653] at PRS020534).

The Content Advisory Committee (the “CAC”) then reviewed the framework for content appropriateness, which included “content accuracy, significance, job-relatedness, and freedom from bias.” (ALST Tech. Manual, at PRS012626). Although the SED states that the CAC consisted of twenty-seven educators in total, only fourteen of them actually participated in the review of the ALST Framework. (*Id.* at PRS012627); (ALST Tech. Manual, App. K, at PRS012922–23). Two of the fourteen participating CAC members identified as African-American, one identified as multiracial, and all others identified as Caucasian. (Pearson Ex. 60, at PRS020534).

After the two committees assessed the Framework, the SED reviewed and then adopted each committee’s recommendations for changes. *See* (ALST Tech. Manual, at PRS012629).

Next, Pearson conducted a “content validation survey.” The survey was distributed to approximately 500 New York State public school teachers, and asked each teacher to rate, on a scale of 1 to 5,¹¹ the following information:

- (1) the importance of the ALST KSAs (“Reading” and “Writing to Sources”) for performing the job of a New York State public school teacher;
- (2) how well each of the Performance Indicators used to define those KSAs represent “important examples” of those KSAs;
- (3) how well the ALST KSAs, as a whole, “represent[] important aspects of the knowledge and skills needed to teach in New York State public schools.”

(*Id.* at PRS012630).

Of the approximately 500 surveys Pearson distributed, 223 were completed and eligible for use in the data analysis. (*Id.* at PRS012633). The demographic characteristics of the respondents were as follows:

- Caucasian: 81.8% (180 respondents);
- African-American: 3.6% (8 respondents);
- Latino: 9.1% (20 respondents);
- Female: 68.9% (153 respondents);
- Male: 31.1% (69 respondents).

(ALST Tech. Manual, App. O [ECF No. 652-7] at PRS013453).

¹¹ Dr. Outtz testified that the better way of determining the importance of KSAs (or job tasks) was to ask each survey respondent to distribute a defined number of points across the applicable KSAs. *See* (June 22, 2015 Hr’g Tr. 40). For example, a respondent might be asked to distribute 100 points in total and might apportion 20 points to one KSA, 30 points to another, and 10 points to another. That respondent would then have only 40 points remaining to distribute to the remaining KSAs. In other words, the more points a respondent apportions to one KSA, the fewer points she has to distribute to the others. (*Id.*) The Court concurs with Dr. Outtz’s testimony. The point-distribution approach allows the respondent to consider more thoughtfully, and to communicate more accurately, the relative importance of those KSAs. (*Id.*) Determining the relative importance of job tasks and KSAs is an important part of the required job analysis. *See M.O.C.H.A. Soc’y, Inc. v. City of Buffalo*, 689 F.3d 263, 278 (2d Cir. 2012).

The same survey was sent to 112 faculty members who teach educator preparation programs in New York State. (ALST Tech. Manual, at PRS012633–36). Sixty-three returned a completed survey eligible for use in the data analysis. (*Id.* at PRS012636). The demographic characteristics of this 63-member sample were as follows:

- Caucasian: 95.2% (59 respondents);
- African-American: 0% (0 respondents);
- Latino: 0% (0 respondents);
- Female: 69.8% (44 respondents);
- Male: 30.2% (19 respondents).

(ALST Tech. Manual, App. O, at PRS013454).

The results of both the public school teacher and the educator preparation faculty surveys showed that respondents viewed the KSAs included in the ALST Framework as of either “great importance” (a rating of 4), or “very great importance” (a rating of 5). (Gullion Decl. ¶¶ 29–30). Respondents rated the Performance Indicators similarly. (*Id.*) Pearson analyzed responses from African-American and Latino respondents separately, and found that both groups rated the KSAs and Performance Indicators similarly to the survey respondents as a whole. *See (id.* ¶ 29); (ALST Tech. Manual, App. O, at PRS013456).

These results led Pearson to conclude that “Reading” and “Writing to sources” constituted important KSAs necessary for the successful performance of a New York State public school teacher’s job. *See* (ALST Tech. Manual, at PRS012638) (“The Content Validation survey results confirm that the set of competencies are important job tasks that all teachers must perform and that the competencies are appropriate to be assessed.”).

ii. The Job Tasks Analysis

At some point after Pearson completed its ALST Framework analysis, Pearson sub-contracted with HumRRO to complete a job tasks analysis. *See* (HumRRO Report, at

PRS013013) (noting that the ALST Framework existed at the time HumRRO began its job tasks analysis). The goal of the job tasks analysis was to compile a list of the job tasks a New York State public school teacher performs as a part of her daily responsibilities, and then to verify that those tasks are important to all of New York State’s public school teachers. (Paullin Decl. ¶¶ 15–17).

A HumRRO “job analysis expert” began by training two Pearson staff members in how to develop a job task list from a list of documents HumRRO deemed relevant to the task. (HumRRO Report, at PRS013020–21). The document list included the Teaching Standards and Common Core Standards, other standards documents developed by state or national organizations, academic articles discussing teaching practices, and an online database of job tasks called the O*NET. (*Id.*); *see also About O*NET*, O*NET Resource Center, <http://www.onetcenter.org/overview.html> (last visited July 9, 2015). The two Pearson staff members drafted an initial list of 101 tasks, divided into seven categories. (HumRRO Report, at PRS013021).

Next, HumRRO assembled a focus group of “subject matter experts” (or “SMEs”), consisting of New York state public school educators and education preparation program faculty. (*Id.* at PRS013022–23). HumRRO referred to this focus group as the “Job Analysis Task Force.” (Paullin Decl. ¶¶ 52–53). Twenty-five SMEs participated in the Task Force; twenty-two of them were Caucasian, and the remaining three failed to identify their race or ethnicity. (Pearson Ex. 60, at PRS020534). No member of the Task Force identified as African-American or Latino. (*Id.*) The Task Force reviewed the initial job task list Pearson drafted, and suggested wording changes, deletions, and additions to the list. (HumRRO Report, at PRS013023–24). The final task list included 105 tasks across 7 categories. (*Id.* at PRS013024).

HumRRO then administered a survey asking New York educators to rate, on a scale of 1 to 5, the importance of each of the 105 job tasks, as well as the frequency with which the educators performed each task. (*Id.* at PRS013025). The survey also asked whether HumRRO’s job task list was comprehensive, and provided an opportunity for respondents to identify critical tasks not included in the task list. (*Id.* at PRS013035).

HumRRO invited 7,033 teachers to complete the survey, and received 1,655 completed surveys that were eligible for data analysis. (*Id.* at PRS013028). The demographic characteristics of the respondents were as follows:

- Caucasian: 84.4% (1397 respondents);
- African-American: 3.0% (49 respondents);
- Latino: 5.5% (92 respondents);
- Female: 73.0% (1,208 respondents);
- Male: 25.5% (422 respondents).¹²

(*Id.* at PRS013030).

HumRRO divided responses into 18 different “assignment groups,” based on the population a teacher taught. Examples of assignment groups included: teachers of grades 1–6, teachers of grades 5–9, social studies teachers, math teachers, teachers of students with disabilities, and teachers of world languages. *See (id.* at PRS013026) (listing all of the 18 assignment groups). Across assignment groups, all of the tasks received a mean importance rating of 3.0 or higher, meaning respondents judged all of the 105 tasks to be at least “important.” (*Id.* at PRS013031–32). Additionally, 86% of respondents stated that the 105 job tasks covered at least 80% of their job; 20% stated, however, that one or more critical tasks were missing from the list. (*Id.* at PRS013035–36).

¹² Two percent of respondents (25 individuals) did not report their gender. (*Id.* at PRS013030).

HumRRO analyzed the additional job tasks that some respondents claimed were critical, but that were missing from the job task list, and determined that most of them either were variations of tasks already included in the list, or were KSAs and therefore inappropriate to include in the job task list. (*Id.* at PRS013036–37). Ten of the tasks respondents listed as missing, however, were not well covered by the existing task list. (Paullin Decl. ¶ 123). HumRRO “recommended that these tasks be added to any future task surveys conducted by Pearson,” but otherwise did nothing to include these tasks in the task list.¹³ (*Id.*)

HumRRO then sought to determine which of the tasks were “critical.” It developed a formula that calculated criticality using a combination of how important teachers rated a task to be, alongside how frequently teachers performed the task. (HumRRO Report, at PRS013033). Relying on this formula, HumRRO determined that 34 of the 105 job tasks should be considered critical. (*Id.*) According to HumRRO’s calculations, these 34 tasks were rated as critical by each of the 18 assignment groups individually, as well as collectively. (*Id.* at PRS013033–35).

The next step in HumRRO’s job task analysis involved linking the job tasks compiled by HumRRO to the KSAs Pearson had defined in its ALST Framework. HumRRO believed that demonstrating such a link would illustrate that the KSAs being tested by the ALST were sufficiently job-related. (*Id.* at PRS013050).

First, HumRRO asked the Job Analysis Task Force, discussed above, to assess the importance of the Performance Indicators used to describe the two KSAs tested by the ALST—“Reading” and “Writing to Sources.” (*Id.* at PRS013045). The Task Force rated the importance

¹³ These tasks were: (1) “Performs administrative tasks,” (2) “Provides support to students and families who are experiencing distress,” (3) “Mentors or provides career counseling to students,” (4) “Manages students who exhibit behavioral problems,” (5) “Attends mandatory meetings,” (6) “Performs extracurricular duties such as coaching teams and clubs, directing musical or artistic performances, serving as a chaperone on class trips,” (7) “Prepares classroom for daily use,” (8) “Grades schoolwork and other assessments,” (9) “Provides extra tutoring or lessons for individual students,” and (10) “Writes letters of recommendation for students.” (Paullin Decl. ¶ 123).

of each Performance Indicator on a scale of 1 to 5. (*Id.*) The group, on average, rated each Performance Indicator as being of either “great importance” (a rating of 4), or “very great importance” (a rating of 5). (HumRRO Report, App. L, [ECF No. 652-6] at PRS013302–05).

Next, a different focus group of SMEs participated in what HumRRO calls a “linkage exercise.” (HumRRO Report, at PRS013050) (internal quotation marks omitted). During this exercise, the SMEs were provided the job tasks list, and were asked to rate, on a scale of 1 to 3, whether the KSA of either “Reading” or “Writing to Sources” was important to that job task. (*Id.* at PRS013051–52). A rating of 1 indicated that the KSA was not important to a given task, while a rating of 2 or 3 indicated that the KSA was either “needed” to perform the task or “essential” to perform the task. (*Id.* at PRS013052). A job task was considered “linked” to a KSA when at least 75% of the SMEs rated the combination as a 2 or 3. (*Id.* at PRS013053). Although HumRRO provided the Performance Indicators to the focus group as a way of defining “Reading” and “Writing to Sources,” *see (id.* at PRS013052), the SMEs were not asked to, and did not, link the Performance Indicators to the job tasks list. *See* (Outtz ALST Report 29–30).

Nine SMEs participated in this focus group; four identified as Caucasian, one identified as African-American, one identified as Asian-American, and three failed to report their race or ethnicity. (Pearson Ex. 60, at PRS020534). No SME in this focus group identified as Latino. (*Id.*)

The focus group found that both “Reading” and “Writing to Sources” were linked to a large number of tasks. The group linked 66 tasks to the KSA of “Reading,” 20 of which were considered “critical”; it linked 61 tasks to the KSA of “Writing to Sources,” 20 of which were considered “critical.” (HumRRO Report, at PRS013054).

By linking the job tasks list developed by HumRRO to the KSAs defined in Pearson’s ALST Framework, HumRRO determined that its job tasks analysis supported the relevance and job-relatedness of the ALST to the job performed by New York State public school teachers. *See (id. at PRS013056–57).*

iii. ALST Test Question Development

The next step in the development process was for Pearson staff members to write the ALST test questions themselves. To facilitate this process, Pearson first created a document it called “assessment specifications,” which provided guidelines concerning how to select the reading passages that test takers would be required to analyze, as well as how to draft the questions that test takers would be asked in response to those reading passages. (ALST Tech. Manual, at PRS012645). The assessment specifications list each Performance Indicator, alongside a more detailed description of how that indicator should be tested. *See generally* (ALST Tech. Manual, App. P [ECF No. 652-8]). Pearson refers to these more detailed descriptions as “Examples of Observable Evidence.” (*Id.*) For instance, for the Performance Indicator “draws conclusions based on textual evidence,” the example of observable evidence states that a candidate should “identif[y] a logical conclusion about the central idea or theme in a text, based on specific textual evidence.” (*Id. at PRS013459*).

Pearson staff members used these assessment specifications to draft the individual ALST test questions. Once written, these questions were reviewed by the BRC and the CAC. (ALST Tech. Manual, at PRS012646–54). These reviews took place in two phases, one occurring in January 2013, and the other in November 2013. *See (id. at PRS012647)* (stating that “[t]he item development effort required to assemble an item bank for ALST was so large that the items were

developed and reviewed in phases”). As a result of these reviews, certain questions were revised, while others were eliminated from testing entirely. *See (id. at PRS012649, PRS012654).*

Although the BRC functioned in the same manner as it had with respect to the development of the ALST Framework, the CAC had an expanded role in this phase of the development process. Members of the CAC reviewed each ALST question by focusing on four concerns:

- (1) Does the question actually measure the KSAs described in the ALST Framework?
- (2) Is the question accurate—in other words, is it factually correct, and is there only one unambiguously correct response?
- (3) Is the question free from bias?
- (4) Does each question measure KSAs that a New York State public school teacher must possess to perform her job competently?

(Id. at PRS012649–53).

Once the BRC and CAC had approved the questions, Pearson field tested them. Field testing involved administering sets of ALST questions to test populations, to determine whether the questions were “appropriate, clear, reasonable, and had acceptable statistical and qualitative characteristics.” *(Id. at PRS012655).*

Pearson used two field testing strategies. First, it field tested stand-alone ALST questions by administering sets of questions to volunteer field testing participants. *(Id.)* These participants were typically individuals who were in the process of completing an educational preparation program. *(Id.)* This was done to ensure that the test population was as similar to the real test taking population as possible. *(Id.)* At least 73 field test participants responded to each multiple-choice question tested during this process. *(Id.)*

Pearson’s second strategy involved the inclusion of “non-scorable” questions on operational (i.e. scored) ALST exams. The multiple choice portion of every ALST exam contains four scorable sets of questions, and one non-scorable set. (Gullion Decl. ¶ 8). The non-

scorable set does not count towards a test taker's score, but the responses to those non-scorable questions are analyzed to ensure that the questions are acceptable to be included on future exams. (ALST Tech. Manual, at PRS012656).

Based on the results of this field testing, Pearson eliminated some test questions and submitted others for revision or further field testing. (Gullion Decl. ¶¶ 61–62).

iv. Determining the Minimum Passing Score

The final step in Pearson's development process was to establish what would constitute a passing score on the ALST. To do so, Pearson convened a focus group of eighteen SMEs to determine what score on the exam "demonstrates the level of performance expected for just acceptably qualified candidates." (ALST Tech. Manual, at PRS012665). The demographic make-up of this focus group was as follows:

- Caucasian: 89% (16 individuals);
- African-American: 11% (2 individuals);
- Latino: 0% (0 individuals);
- Female: 72% (13 individuals);
- Male 28% (5 individuals).

(Pearson Ex. 60, at PRS020535).

Pearson used a "modified Angoff" method and an "extended-Angoff method" to determine the passing score. (ALST Tech. Manual, at PRS012666). These methods involved asking the group to conceptualize "[a] hypothetical individual who is just at the minimum level of academic literacy skills a teacher needs in order to be competent in the classroom and positively contribute to student learning." (*Id.* at PRS012668). The SMEs were asked to determine what percentage of the exam questions this hypothetical individual would answer correctly, and what score he or she would achieve on the essays. (*Id.* at PRS012668–69). Pearson then calculated the focus group's recommended passing score, and provided this

information to the Commissioner of Education, who, in coordination with the SED, determined what the ALST's passing score would be. (*Id.* at PRS012672–73).

V. ANALYSIS

The first step in determining whether the ALST is discriminatory under Title VII is to decide whether the test has a “disparate impact on the basis of race, color, religion, sex, or national origin.” *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009) (quoting 42 U.S.C. § 2000e-2(k)(1)(A)(i)). The parties disagree as to whether a disparate impact exists here.¹⁴ The Court does not need to decide the issue, however, because the Court holds that the ALST is job related. This is sufficient to rebut any showing of disparate impact Plaintiffs might have made. *See N.A.A.C.P., Inc.*, 70 F.3d at 225.

A. The SED and Pearson Properly Relied on the Standards Documents

To determine whether an exam was properly validated as job related by its designers, courts must first decide whether the designers appropriately ascertained the content of the job.

¹⁴ The dispute concerns which set of test-takers is the appropriate one to count for the purposes of disparate impact analysis: Plaintiffs and Dr. Outtz believe that all test-takers should count, *see* (Plaintiffs July 2, 2015 Ltr. [ECF No. 624] at 2–3); (June 22, 2015 Hr’g Tr. 51); the SED and Pearson believe that the only test-takers who should count are those who have completed an education preparation program funded by Title II, *see* (SED Post-Trial Br. [ECF No. 648] at 2–4); (Karlin Decl. [ECF No. 641] at ¶¶ 34–39). The SED and Pearson refer to these individuals as “program completers.” *See* (SED Post-Trial Br. 2–4); (Karlin Decl. ¶¶ 34–39). Plaintiffs and Dr. Outtz argue that because the SED allows anyone to take the ALST, all test-takers should be included for disparate impact analysis purposes. (Plaintiffs July 2, 2015 Ltr. 2–3); (June 22, 2015 Hr’g Tr. 51). The SED and Pearson contend that because program completers “are the only candidates who might be denied a job if failing the ALST is the only reason that they are ineligible for New York educator certification,” such program completers should be the only individuals counted. (Karlin Decl. ¶ 36); *accord* (SED Post-Trial Br. at 2–4).

Plaintiffs and Dr. Outtz assert that Plaintiffs have made a *prima facie* showing of discrimination because when taking the pass rates of all test-takers into account, African-American and Latino test-takers fall well below the 80% threshold that typically constitutes a disparate impact. *See* (Outtz ALST Report 7–8); *see also* Guidelines § 1607.4(D); *Vulcan Soc’y*, 637 F. Supp. 2d at 87. The SED and Pearson disagree, and contend that when only program completers are considered in the population being assessed for disparate impact, no disparate impact sufficient to make a *prima facie* showing of discrimination exists. *See* (SED Post-Trial Br. 2–4); (Karlin Decl. ¶¶ 34–39). Because the Court does not need to decide whether Plaintiffs have made a *prima facie* showing of disparate impact, it does not need to decide whether disparate impact should be calculated using only program completers, or if the general population should be used instead.

See Albemarle Paper Co., 422 U.S. at 431 (stating that a court must decide whether an exam is “predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.” (internal quotation marks omitted)). In most cases, the best way for test designers to ascertain the content of a job is to first gather information about the job directly from job incumbents. *See M.O.C.H.A. Soc’y*, 689 F.3d at 278 (noting that “a proper job analysis . . . is conducted by interviewing workers, supervisors and administrators; consulting training manuals; and closely observing the actual performance of the job” (internal quotation marks omitted)); *see also id.* (stating that “[a]lthough interviews and observations may be best practices,” they are not “an absolute requirement for every job analysis,” and that “[i]n some circumstances, it may be possible to gather reliable job-specific information by other means, such as survey instruments sent directly to employees”). This is the function that the job analysis plays in the validation process; it is used to gather information about the tasks and skills that make up the job. *See Guidelines § 1607.14(C)(2)* (“There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance Any job analysis should focus on the work behavior(s) and the tasks associated with them.”).

Written job descriptions can also be useful to determining the content of a job, particularly when those descriptions were promulgated by an organization or government agency with expertise in the field. *See Ass’n of Mexican-Am. Educators v. California*, 937 F. Supp. 1397, 1418 (N.D. Cal. 1996) (finding a “manifest relationship” between the skills on the exam in question and the job of teaching based on, *inter alia*, standards promulgated by national teaching organizations); *see also Ass’n of Mexican-Am. Educators v. California*, 231 F.3d 572, 589 (9th

Cir. 2000) (finding the fact that “the kinds of skills tested in the [certification exam in question] can be found in elementary and secondary school textbooks” lends credence to the validity of the exam). However, written descriptions are typically insufficient by themselves because they do not provide as useful or accurate information as do job incumbents. *Cf. M.O.C.H.A. Soc’y*, 689 F.3d at 278. Written job descriptions may not capture the nuances or totality of an occupation, particularly not one as complex as teaching. They also may be outdated and may not take into account new approaches to the job.¹⁵

Nonetheless, in the unique circumstances of the ALST’s creation, the written job descriptions here—the Teaching Standards and the Common Core Standards (collectively, “the Standards”)—constituted a sufficient basis for ascertaining critically important aspects of a New York State public school teacher’s job. At the time that the SED and Pearson designed the ALST, New York was just beginning to implement the comprehensive educational reforms that had secured New York access to federal funding under the Race to the Top initiative. To develop the ALST, the SED and Pearson began not with an analysis of what teachers *now* do on the job, but an analysis of the KSAs a teacher’s job *will* require going forward. This was necessary because the ALST was designed to reflect these new educational reforms. (*Id.*) The Standards, which stemmed directly from those reforms, do not just *describe* the job of teaching, they *transform* it. *See* (Wagner Decl. ¶ 18) (asserting that New York’s teaching reforms “aimed at transforming teaching and learning in New York State”). These Standards may not encompass every aspect of a teacher’s job, but they encompass enough critical aspects of the job to justify

¹⁵ Misplaced reliance on descriptive documents in connection with a job analysis played a significant role in this Court’s decision holding the LAST-2 improperly validated. *See Gulino V*, 2015 WL 3536694, at *12, *16. In that decision, the Court found that the SED determined the KSAs assessed by the LAST-2 based on documents “describing common liberal arts and science course requirements at New York state colleges and universities,” *id.* at *7, *12, which were inadequate for the reasons explained in that Opinion and discussed further below.

testing to determine whether applicants possess the skills necessary to perform competently these new portions of the job. *See infra* Part V.D (holding that the ALST is representative of a teacher’s job because the literacy skills it tests are more than a minor part of the skills the job requires).

By describing what and how New York’s students must be taught, the Standards provide the specific information the test developers needed to determine what these new portions of a teacher’s job were going to entail, based on the Race to the Top reforms, and how the KSAs necessary to perform this new incarnation of the job could be tested.¹⁶ Because New York State implemented its Race to the Top reforms so close in time to the development of the ALST—in fact, the development of the ALST was a part of the reform process itself—job incumbents could not have been expected to know the information Pearson needed to design a test that assessed the new sets of skills New York State expected its teachers to possess once these reforms were adopted.¹⁷

The Standards were a particularly sound basis for validation because they contain so much detail about how and what teachers should teach. For example, the Common Core

¹⁶ It is for this reason that the Court respectfully disagrees with Dr. Outtz’s conclusion that the SED failed to demonstrate that the ALST was valid. *See* (Outtz ALST Report 32–33). Dr. Outtz’s thoughtful and rigorous report highlighted many of the problems that exist in Pearson’s and HumRRO’s test development process. However, most of those problems relate to the job analysis. Although the Court agrees that Pearson’s and HumRRO’s job analysis was imperfect, *see infra* Part V.G, those problems with the job analysis are not relevant to the Court’s decision here. Pearson’s use of the Standards to define the job of teaching suffices as a job analysis.

¹⁷ The Court pauses here to highlight just how much control these Standards indicate the SED has over the daily responsibilities of teachers. It is not just curriculum (although that too appears to be controlled by the SED, as is clear from their implementation of the Common Core Standards, *see generally* (ALST Tech. Manual, App. L)), or certification that the SED regulates, but the intricacies of a teacher’s daily job tasks. The Teaching Standards define how teachers are supposed to interact with students, *see* (ALST Tech. Manual, App. A, at PRS012706–07), how they are to construct their lesson plans, *see (id.* at PRS012708–12), the environment that teachers are to facilitate among their students, *see (id.* at PRS012713–14), how teachers are to assess their students’ achievement, *see (id.* at PRS012715–16), how teachers are to interact with their colleagues, *see (id.* at PRS012717–18), and the ways teachers should continue their own learning, *see (id.* at PRS012719–20). This level of control did not exist when the Second Circuit decided *Gulino II*. *See Gulino II*, 460 F.3d at 379 (stating that the only control the SED has over public school teachers is that “it controls basic curriculum and credentialing requirements”). Such a stark increase in the SED’s regulatory control over teachers may warrant revisiting the portion of *Gulino II* that held that the SED was not an employer of New York State’s public school teachers. *See infra* note 25.

Standards define with great specificity the degree of literacy that teachers are now expected to instill in their students. The Performance Indicators contained within the ALST Framework almost mirror these Common Core Standards.¹⁸ The following grid provides an illustration:

Performance Indicator	Common Core Standard
Teachers must be able to “determine[] what a text says explicitly.” (ALST Tech. Manual, App. I, at PRS012914).	Students must be able to “[c]ite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.” (ALST Tech. Manual, App. L, at PRS012975).
Teachers must be able to “make[] logical inferences based on textual evidence.” (ALST Tech. Manual, App. I, at PRS012914).	
Teachers must be able to “analyze[] the development of central ideas or themes of a text.” (ALST Tech. Manual, App. I, at PRS012915).	Students must be able to “[d]etermine two or more themes or central ideas of a text and analyze their development over the course of the text, including how they interact and build on one another to produce a complex account.” (ALST Tech. Manual, App. L, at PRS012975).
Teachers must be able to “analyze[] how and why individuals, events, and ideas develop and interact over the course of a text.” (ALST Tech. Manual, App. I, at PRS012915).	

Kenneth Wagner, a deputy commissioner for the SED, credibly testified that all teachers—whether they teach, for example, English, Math, Biology, or Art—are required to integrate into their curriculum the literacy skills identified by the Common Core Standards. (June 22, 2015 Hr’g Tr. 108). It is not unreasonable for New York State to require each of those

¹⁸ Pearson testified that the ALST Framework is based upon the portion of the Common Core Standards intended for eleventh grade students. (June 22, 2015 Hr’g Tr. 161).

teachers to demonstrate fluency in the very literacy skills that they are required to teach to their students. *See* (Wagner Decl. ¶¶ 49–50).

Plaintiffs argue that these Standards are no different from the documents from which NES derived the LAST-2, which the Court held to be an insufficient basis for exam validation in *Gulino V.* *See* (July 20, 2015 Hr’g Tr. [ECF No. 656] at 4–5). Plaintiffs are incorrect; these Standards differ in meaningful ways from the documents at issue in *Gulino V.* To develop the LAST-2, NES relied primarily upon documents that described “common liberal arts and science course requirements at New York state colleges and universities,” *Gulino V.*, 2015 WL 3536694, at *7, as well as “syllabi and course outlines for courses used to satisfy those liberal arts and sciences requirements,” (Clayton Decl. ¶ 14). These documents are clearly less descriptive of the job of teaching, and therefore are inferior to the Standards at issue here. They did not describe the job of teaching in any direct way; they simply described how the liberal arts (the skill purportedly tested by the LAST-2) were taught to educators-in-training.

With respect to the LAST-2, NES also claimed to have relied on “numerous materials that define and describe the job of a New York State teacher, including New York State regulations and guidelines for teachers, student learning standards, textbooks and other curricular materials.” (*Id.* ¶ 15). However, none of these materials was ever provided to the Court, and therefore, the Court was not able to appraise the value of these documents. Moreover, there was no indication that these documents reflected a transformation of the job of teacher. In other words, these documents would not have described the job of teaching any more accurately or comprehensively than job incumbents would have done, had they been questioned. As the Court noted above, documents such as these are likely inferior to the information that job incumbents could provide, due to the documents being incomplete, unclear, or outdated.

Accordingly, the Court finds that the Standards appropriately and sufficiently establish what New York State expects the transformed job of its public school teachers to entail, which would not have been reflected in a traditional job analysis.¹⁹ The SED's and Pearson's reliance

¹⁹ Plaintiffs assert that the Standards cannot relieve the SED and Pearson of its duty to conduct a job analysis simply because New York State has changed its approach to the job of teaching. (July 20, 2015 Hr'g Tr. 4–7). They contend that experts in the test development field describe the job analyses required in such instances as “futures job analyses.” (*Id.* at 20–21). According to Plaintiffs, a futures job analysis requires greater rigor than a normal job analysis, and Plaintiffs contend that the SED and Pearson have failed to demonstrate such rigor with respect to the ALST. (*Id.*) The Court has not found sufficient support for Plaintiffs' contention.

Plaintiffs submitted two articles by test development experts, as well as a PowerPoint presentation, that describe futures job analyses. *See* (July 21, 2015 Ltr. [ECF No. 647]). The first article Plaintiffs provided—*Strategic Job Analysis*—describes potential methods for conducting a job analysis that take into account how a job *might* change in the future. *See generally* Benjamin Schneider & Andrea Marcus Konz, *Strategic Job Analysis*, 28 Hum. Resource Mgmt. 51 (Spring 1989). These methods ask job incumbents and other SMEs to brainstorm about what those future changes might look like and how the job tasks required by the job might change accordingly. *Id.* at 54–56. The methods recounted in the article do not apply here, because here there is no need to brainstorm about how the job of teaching *might* change in the future. The Standards specifically communicate the way New York State expected the job of teaching to change, and the SED sought to validate an exam that would test for some of the skills required to successfully perform this newly-changed job. There was no need for a job analysis here that asked job incumbents and other SMEs how the job *might* change in the future.

The PowerPoint presentation similarly addresses an organization that is considering the effects of potential future job changes, not one where those changes have already been determined and implemented. *See generally* Ilene Gast & Kathlea Vaughn, *Strategic Job Analysis: Predicting Future Job Requirements* (2012), <http://www.ipacweb.org/Resources/Documents/conf12/gast.pdf>.

The final article—*Future-Oriented Job Analysis: A Description of the Process and Its Organizational Implications*—discusses a futures job analysis test developers applied to an organization that was reorganizing by merging eleven old positions into three new positions. *See* Ronald S. Landis et al., *Future-Oriented Job Analysis: A Description of the Process and Its Organizational Implications*, 6 Int'l J. of Selection and Assessment 192, 192 (July 1998). The test developers there performed a futures job analysis to determine the content of these three new positions and how to test for the skills these new positions would require. *Id.* at 193–95. Similar to the other article Plaintiffs provided, there was some uncertainty about how these new positions would function. *See id.* at 193 (stating that there was “some ambiguity about the allocation of tasks to the three new positions,” and that “different perspectives regarding the jobs existed.”). The futures job analysis served as a way of concretizing how these new positions would be performed. *Id.* (noting that the job analysis “allowed the consulting team to develop a comprehensive understanding of which critical tasks were reallocated under the new organizational structure, which were eliminated or modified due to new information systems, and which were unaccounted for in the new design”).

Additionally, much of the job analysis described in the Landis article involved the sorting of tasks. *Id.* at 194–95. It appears that most of the tasks that were to be performed by the three new positions were already performed by the eleven positions that were being phased out. *See id.* at 193 (“Many of the tasks expected to be part of the new jobs were currently being performed across the existing 11 positions.”). Thus, job incumbents had valuable information about how those tasks were currently performed; that information was taken into account when determining which of these tasks would be performed by which of the new positions. *Id.* at 195.

The methods described in the Landis article also would not be appropriate in this instance. Here, the tasks that teachers are to perform have already been set forth in the Standards; the kind of task sorting noted in the article would not have been appropriate here, nor is there confusion about what the position would look like that a futures job analysis could clarify. Moreover, the KSAs tested by the ALST are directly linked to the Common Core Standards through the Performance Indicators, which clearly describe the literacy skills teachers are expected to instill in their students. A job analysis to determine what teacher tasks are going to exist in the future, or what skills are going to be needed to perform them, would not be useful, because the skills needed are already defined in the

on the Standards is sufficient to comply with the first *Guardians* factor—that “test-makers must have conducted a suitable job analysis,” *Guardians*, 630 F.2d at 95. The SED must also establish that the ALST tests for critical aspects of the job as the Standards define it. To do so, it must demonstrate compliance with the remainder of the *Guardians* factors. The SED has done so here.

B. Pearson Used Reasonable Competence in Constructing the ALST

The second *Guardians* factor requires an examination of whether the employer “used reasonable competence in constructing the test.” *Guardians*, 630 F.2d at 95. Test developers are generally viewed as having used reasonable competence if the exam was created by professional test preparers, and if a sample study was performed that “ensure[d] that the questions were comprehensible and unambiguous.” *M.O.C.H.A. Soc’y*, 689 F.3d at 280. Here, Pearson and HumRRO are both professional test preparers. (Gullion Decl. ¶ 3); (Paullin ¶¶ 7–10). Pearson field tested all of the ALST questions to ensure that they were comprehensible and unambiguous. (ALST Tech. Manual, at PRS012655–61). Accordingly, it is clear that Pearson used reasonable competence in constructing the ALST.

C. The Content of the ALST is Related to the Job of Teaching

The third *Guardians* factor requires the content of the exam to be directly related to the content of the job. This requirement “reflects ‘[t]he central requirement of Title VII’ that a test be job-related.” *Vulcan Soc’y*, 637 F. Supp. 2d at 116 (quoting *Guardians*, 630 F.2d at 97–98).

The Court credits Mr. Wagner’s testimony that literacy skills, as defined by the Common Core Standards, are a critical component of the skills that teachers are required to teach their

Common Core Standards. Any futures job analysis to determine those skills at this point in time would be superfluous.

students. *See* (Wagner Decl. ¶¶ 47–51). An exam that tests for the literacy skills that a teacher must instill in her students is inherently job related. Therefore, if the SED has demonstrated that the ALST tests for the literacy skills set forth in the Common Core Standards, the SED has shown that the ALST is job related, and therefore, is not discriminatory under Title VII.

The ALST Framework that Pearson devised identified two KSAs—“Reading” and “Writing to Sources.” Contained within each of those KSAs are a number of Performance Indicators, which Pearson used to explain in more detail the nature and range of the knowledge and skills encompassed by these two KSAs. As noted above, these Performance Indicators nearly mirror the Common Core Standards. *See supra* Part V.A. These Performance Indicators were used appropriately to devise the test questions included in the ALST.

Pearson used the Performance Indicators to create the “assessment specifications” that the test question writers relied upon to formulate the exam’s test questions. (June 22, 2015 Hr’g Tr. 153–56); (ALST Tech. Manual, App. P). The assessment specifications expanded on each Performance Indicator, giving test question writers the detailed information they needed to ensure that the test questions meaningfully tested for the skills described by those Performance Indicators. (June 22, 2015 Hr’g Tr. 153–56); (ALST Tech. Manual, App. P). These test questions were then reviewed by the BRC and the CAC. (ALST Tech. Manual, at PRS012646–54). The CAC was specifically tasked with confirming that the test questions actually measured the KSAs listed in the ALST Framework—the Framework that was shaped by the Standards. (*Id.* at PRS012651). The test questions were then field tested rigorously to establish that the questions and answers were clear, accurate, and unambiguous. (*Id.* at PRS012655–61). These procedures are sufficient to demonstrate that the Performance Indicators were used appropriately to devise the ALST’s test questions.

Accordingly, the Court holds that the content of the ALST is related to the job of teaching in New York State public schools.

D. The Content of the ALST is Representative of the Content of a Teacher’s Job

The fourth *Guardians* requirement is that the content of the exam must be “a representative sample of the content of the job.” *Guardians*, 630 F.2d at 98 (internal quotation marks omitted). This does not mean that “all the knowledges, skills, or abilities required for the job [must] be tested for, each in its proper proportion.” *Guardians*, 630 F.2d at 98. Rather, this requirement is meant to ensure that the exam “measures important aspects of the job, and does not overemphasize minor aspects.” *Gulino III*, 907 F. Supp. 2d at 521 (citing *Guardians*, 630 F.2d at 98). Here, the literacy skills tested by the ALST are not a minor aspect of the job. Pearson’s validation process linked 20 of the 34 critical tasks, and between 61 and 66 tasks overall, to the KSAs of “Reading” and “Writing to Sources”—KSAs elaborated upon by Performance Indicators that closely correlated to the Common Core Standards. (HumRRO Report, at PRS013050–55). In other words, the teachers surveyed by Pearson found Common Core literacy skills to be important to more than half of a teacher’s daily job tasks.²⁰

²⁰ As the Court notes below, *see infra* Part V.G.ii, HumRRO should have linked job tasks to what Pearson referred to as the “Performance Indicators,” and not just what Pearson termed the “KSAs.” That is because it is the Performance Indicators that actually represent the skills tested by the ALST. Nonetheless, the fact that a focus group of SMEs linked Pearson’s two KSAs, as elaborated upon by the Performance Indicators, to HumRRO’s list of job tasks is sufficient to meet the requirements of the fourth *Guardians* factor. *Contrast Vulcan Soc’y*, 637 F. Supp. 3d at 119–20 (finding the exam there unrepresentative of the content of the job after determining that the job included “a host of requirements other than cognitive abilities,” and that only one of the 21 “job clusters” identified by the test proponents in the job analysis required cognitive abilities, yet noting that the exam in question tested only cognitive abilities). *Cf. Guardians*, 630 F.2d at 99 (holding the exam there to be representative of the content of the job, despite the fact that “this conclusion would have been easier to reach if the City had spelled out the relationship between the abilities that were tested for and the job behaviors that had been identified,” and noting that “the relationship is sufficiently apparent to indicate that the [test proponent] was not seizing on minor aspects of the . . . job”).

Accordingly, the Court holds that the content of the ALST is representative of the content of a New York State public school teacher's job.

E. The ALST's Scoring System Usefully Selects Those Applicants Who Can Better Perform the Job

In the fifth and final step of the *Guardians* test, a court must determine whether the exam is scored in a way that usefully selects those applicants who can better perform the job.

Guardians, 630 F.2d at 105. A minimum passing score must be set "so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force." *Id.*

(quoting Guidelines § 1607.5(H)). An employer can set the minimum passing score on the basis of "a professional estimate of the requisite ability levels," or "by analyzing the test results to locate a logical 'break-point' in the distribution of scores." *Id.* To establish that a minimum passing score is valid, an employer must present evidence that the score measures the minimum qualifications necessary to succeed at the job. *See Vulcan Soc'y*, 637 F. Supp. 2d at 125.

Here, Pearson used the Modified Angoff method, which is an accepted method in the field for determining a minimum passing score. *See* (Oultz LAST-2 Report 39–41); *Gulino I*, 2003 WL 25764041, at *16 (noting that the Angoff Method "is the most commonly used method for setting a passing score and meets generally accepted professional standards"); *Ass'n of Mexican-Am. Educators*, 937 F. Supp. at 1423–26 (holding that passing scores determined through use of the Angoff Method were acceptable). The Modified Angoff method asks participants in a standard setting focus group to imagine a hypothetical applicant who possesses the minimally adequate skills required for a job, and then decide how that person would perform on the test. (ALST Tech. Manual, at PRS12668–69). The SMEs did that in this case, and the passing score was set accordingly. (*Id.*) This procedure meets the requirements of *Guardians*,

and the Court therefore holds that the ALST's scoring system usefully selects those applicants who can better perform the job of a New York State public school teacher.

F. Pearson's Development of the ALST Complies With the Necessary *Guardians* Factors

In sum, the ALST complies with all of the *Guardians* factors. The SED does not need to demonstrate that it complied with the first *Guardians* factor in the traditional way, because its reliance on the Standards in this instance is an appropriate surrogate for a job analysis in the unique circumstances of this case. Accordingly, the Court holds that the ALST is job related. Defendants have therefore rebutted any *prima facie* case of discrimination made by Plaintiffs. It follows that the ALST is not discriminatory under Title VII.²¹

However, the Court cautions the parties that the situation here may well be *sui generis*. The only reason a traditional job analysis was not required here is because New York State created such all-encompassing new Standards to establish what a teaching job entails, and because the development of the SED's certification regime followed so quickly on the heels of the Standards' development, such that incumbents were not a particularly useful source for determining what skills the job would require. In most challenges to an employment exam, standards documents such as the ones here will not exist; even if they do, some standards may not be sufficiently comprehensive or detailed, such that reliance on those standards alone would be insufficient to determine the content of the job in question. Moreover, in many cases, the test in question will have been devised long enough after the creation of the relevant standards that those standards would have been fully assimilated into the job in question. In such a case, job

²¹ Plaintiffs have not claimed that there was any pretext for actual discrimination here, and therefore the Court's finding that the ALST is job related is dispositive.

incumbents would likely be able to provide information about what the job entails that would be at least as accurate as standards documents, and perhaps more so.

Thus, in most cases, reliance on documents that attempt to define a job, without acquiring information about the content of the job directly from job incumbents, will not be an acceptable means of developing an employment exam, and will not substitute for a properly performed, traditional job analysis.

G. The Flaws in the ALST Development Process

As Dr. Outtz points out in his report, *see* (Outtz ALST Report 11–32), there are several flaws in the way the SED, Pearson, and HumRRO developed the ALST. Had Pearson been unable to rely upon the Standards to understand fully the job of New York State public school teachers, those flaws might have led the Court to find the ALST invalid. Because it may prove helpful in the future, the Court will mention some of these flaws in the remainder of this Opinion.

The SED, Pearson, and HumRRO committed two main errors in the development and attempted validation of the ALST. First, Pearson’s and HumRRO’s procedures were insufficient to ensure that the educators who participated in the development and validation process were sufficiently representative of the New York State public school teacher population. Second, HumRRO failed to link job tasks to the true ALST KSAs—those skills Pearson describes as “Performance Indicators.”

i. Careful Attention Should Be Paid to Ensuring that Focus Groups and Survey Samples Are Representative

Pearson and HumRRO should have done more to ensure that the focus groups and survey samples it used were sufficiently representative. For example, HumRRO’s job analysis task

force, which was charged with developing a comprehensive list of all the job tasks teachers across New York State perform day-to-day, consisted of twenty-five teachers, none of whom identified as African-American or Latino. Similarly, Pearson used two separate committees to assess the ALST Framework it had drafted. Of the fourteen SMEs who served on the CRC, only two identified as African-American, and none identified as Latino.²² (Pearson Ex. 60, at PRS020534). The BRC, which was composed entirely of teachers of color, had only four SMEs, two of whom identified as African-American, and one of whom identified as Latino. (*Id.*) Thus, the views of only three African-American teachers and one Latino teacher were taken into account in the entire drafting of the ALST framework.²³

²² One committee member identified as “multiracial.” (Pearson Ex. 60, at PRS020534).

²³ Several other unrepresentative procedures deserve mention. HumRRO’s job linkage panel consisted of nine SMEs, only one of whom identified as African-American, and none of whom identified as Latino. (Pearson Ex. 60, at PRS020534). The focus group Pearson used to set the ALST’s passing score consisted of eighteen SMEs, two of whom identified as African-American, and none of whom identified as Latino. (*Id.* at PRS020535).

Additionally, Pearson sent out its content validity survey to 500 educators, and received responses from 223 of them. (ALST Tech. Manual, at PRS012633). Of those 223 respondents, only 8 (3.6%) were African-American and only 20 (9.1%) were Latino. (Pearson Ex. 60, at PRS020534). Although the percentages of African-American and Latino responses are similar to the percentages of African-American and Latino teachers in the population as a whole, *see Gulino V*, 2015 WL 3536694, at *13 (stating that approximately 5% of New York’s public school teachers are African American, and 7% are Latino), the raw numbers themselves are unacceptably low. The portion of the content validity survey that was sent to education preparation faculty also yielded unrepresentative results. Of the 63 respondents who completed the survey, none of them identified as African-American or Latino. (Pearson Ex. 60, at PRS020534).

HumRRO’s job analysis survey provides a useful point of comparison. HumRRO sent its survey out to over 7,000 teachers—fourteen times as many as Pearson’s survey—and received seven times as many responses as Pearson. (HumRRO Report, at PRS013028–30). In doing so, HumRRO was able to analyze whether African-American and Latino responses differed meaningfully from the survey respondents as a whole because, in raw numbers, HumRRO received a far greater number of responses from African-American and Latino respondents. (*Id.* at PRS013029–30) (stating that 49 respondents (3%) identified as African-American, and 92 respondents (6%) identified as Latino); (Paullin Decl. ¶ 113) (noting that HumRRO separately analyzed the responses of African-American and Latino teachers, and compared them to the responses of respondents as a whole). The Court is not aware of any reason why Pearson could not have sent out its content validity survey to as many teachers as did HumRRO. Had it done so, Pearson would have received a far greater number of responses overall, and in turn, a likely far greater number of responses from African-American and Latino teachers.

Pearson did compare the responses of African-American and Latino respondents to the respondents as a whole, and found no significant difference in responses. *See* (ALST Tech. Manual, App. O, at PRS013456). This was a laudable effort to analyze data in search of potential bias. Nonetheless, the comparison between eight African-American respondents and the rest of the survey population is not particularly useful or statistically meaningful. According to Pearson, a minimum of 25 respondents for each respondent group is necessary in order to compare the responses of different groups in a meaningful way. *See* (ALST Tech. Manual, at PRS012637). It is unclear, then, why Pearson decided it could compare the responses of African-American and Latino respondents to the respondents as a whole, because only 8 responses were received from African-American teachers and only 20

In all of these instances, Pearson or HumRRO should have taken greater care to ensure that these focus groups were appropriately representative. To perform a proper job analysis, Pearson and HumRRO needed to gather information from job incumbents representative of *all* New York State public school teachers, not just a subset of those teachers. *See Gulino V*, 2015 WL 3536694, at *13–14 & n.21. This does not necessarily mean that different demographic groups must be proportionally represented with statistical accuracy. Particularly with focus groups as small as fifteen or twenty individuals, it might not be possible to achieve group participation in exact proportion to the population at large. Nonetheless, Pearson and HumRRO needed to ensure that all of New York’s public school teachers were represented at more than just a token level.

The SED claims that Pearson and HumRRO each made efforts to recruit African-American and Latino teachers to participate in the development and validation process, (June 22, 2015 Hr’g Tr. 178); (Katz Decl. [ECF No. 644] at ¶¶ 3–8), but initial efforts at recruitment are not enough where they are unsuccessful. If initial recruitment efforts fail, it is inappropriate for a test developer or validator to continue with an unrepresentative group nonetheless.

Although the SED’s and Pearson’s reliance on the Standards is sufficient to save the ALST in this instance,²⁴ unrepresentative data could well contribute to the invalidation of the tests these organizations develop in the future. Accordingly, the Court recommends that the

responses were received from Latino teachers. This low a participation rate by African-American and Latino teachers suggests that any comparison Pearson did between those groups and the respondents as a whole did not yield reliable conclusions. Such conclusions would have been far more reliable if Pearson had surveyed as many teachers in its survey as HumRRO did in its survey. Alternatively, after seeing how few responses it received from African-American and Latino respondents, Pearson could have sent out a second round of surveys targeting these groups.

²⁴ Most of the representativeness issues occurred during the job analysis, where Pearson and HumRRO sought to determine the tasks and skills that define the job of teaching in New York State. Because the Standards allowed Pearson to make those determinations without a traditional job analysis, the lack of representativeness during the job analysis performed here is not fatal.

parties involved in the ALST's development do more in the future to ensure reasonable representativeness in the construction of future survey samples, focus groups, and other reviewing committees.

ii. Job Tasks Must Be Linked to the Actual KSAs Tested by the Exam

Although HumRRO linked the two KSAs listed in the ALST Framework—"Reading" and "Writing to Sources"—to a list of job tasks, it did not do so for the Performance Indicators Pearson used to elaborate upon those KSAs. HumRRO testified that doing so would have required too much work. (Way Decl. [ECF No. 643] at ¶ 31). However, the Court finds Pearson's categorization of KSAs and Performance Indicators flawed. Skills like "reading" and "writing" are overly broad, and do not meaningfully describe the skills being tested by the ALST. Describing an exam as testing "reading" and "writing" says about as much about the content of the exam as saying the exam tests "thinking." Here, it is the Performance Indicators that provide concrete information about what the ALST seeks to assess. "Reading" and "Writing to Sources" are simply convenient means of categorizing the Performance Indicators into comprehensible groups. In this sense, these Performance Indicators are the true KSAs tested by the ALST, and thus, it is the Performance Indicators that HumRRO should have linked to its list of job tasks.

HumRRO did provide the Performance Indicators to the SMEs who participated in the linkage exercise, as a way of elaborating upon what it meant by "Reading" and "Writing to Sources." (Way Decl. ¶ 31). Thus, the SMEs likely took the Performance Indicators into account as they sought to link the skills of "Reading" and "Writing to Sources" to HumRRO's job task list. This suggests that the SMEs believed that at least some of these Performance Indicators were linked to the listed job tasks. However, it is impossible to know which of these

Performance Indicators were so linked; the SMEs may have believed some of them were linked and others were not. Yet, Pearson used all of the Performance Indicators in the assessment specifications the test questions writers relied on to formulate test questions. *See* (ALST Tech. Manual, App. P).

This error is not fatal here, because the Performance Indicators are clearly linked to the Common Core Standards, and that linkage is sufficient, in this instance, to ensure that the Performance Indicators are job-related. In the future, however, the SED (and any subcontractors it hires) will need to be more careful about delineating what is truly a KSA, and what merely constitutes a description of that KSA. When in doubt, test developers and validators should link even those portions of a framework that describe the KSAs (which, in this instance, Pearson terms “Performance Indicators”), to the job tasks, to ensure that the skills actually tested by an exam are linked to the daily tasks an employee performs.

VI. CONCLUSION

For the reasons set forth above, the Court holds that the BOE did not violate Title VII by complying with the SED’s requirement that teachers pass the ALST before they are eligible for employment.²⁵

²⁵ The Court notes that the Second Circuit’s decision in *Gulino II* to dismiss the SED from this lawsuit, although binding precedent that this Court must follow, may be a decision worth revisiting.

The *Gulino II* court remarked that the SED’s dismissal from the case created a “difficult situation” for the BOE, 460 F.3d at 381, but did not discuss or analyze the enormous inequity that the decision created. The Court pauses here to describe that inequity in more depth.

The BOE had no role in the decision to develop and implement state licensing tests such as the LAST-1, the LAST-2, and the ALST, nor did it have any role in the development of any of those exams. *See Gulino V*, 2015 WL 3536694, at *3, 6–8; *Gulino I*, 2003 WL 25764041, at 20–26; *supra* Part IV. To the Court’s knowledge, the SED did not provide any information to the BOE about the processes used and decisions made in connection with the exams’ development—information critical to determining whether the exams were appropriately validated—outside of this class action—long after the tests were used by the SED. Thus, it would appear that the BOE had no way of determining, when the tests were being used by the SED, whether any of these tests were properly or improperly validated, and therefore, whether they were discriminatory. Even if the BOE had been given any of this information, it would have been difficult for it to determine whether the exams were actually discriminatory. Discovery regarding the SED’s development of the LAST-1 took approximately six years, *see* (March 25, 2002

SO ORDERED.

Dated: New York, New York
August 7, 2015

/s/

KIMBA M. WOOD
United States District Judge

Order [ECF No. 59]), and, after a two-month trial, with massive expert testimony, Judge Motely held that the LAST-1 was *not* discriminatory. *See generally Gulino I*, 2003 WL 25764041.

Even if the BOE had been able to determine for itself that the SED's reliance on any of these exams was discriminatory, the BOE would have had little choice but to comply with the SED's requirement that the BOE hire only those teachers who have passed all of the SED-created licensing exams. Failure to do so might have resulted in New York City losing approximately \$7.5 billion dollars' worth of education funding a year—a result that would have crippled New York City's ability to educate its children. *See Gulino V*, 2015 WL 3536694, at *2.

Despite the BOE's complete lack of knowledge concerning the creation or implementation of these exams during the time they were used by the SED, and despite the fact that it had little choice but to comply with the SED's requirement that these exams be used to determine who may and may not be hired to teach, it is the BOE, and not the SED, who remains as the only party liable for the very substantial amount of money being paid to class members.

In the somewhat unique situation posed by this case, where the inequity caused by dismissing the SED from this suit is so great, it could be worth reexamining whether a limited adoption of the “interference test” discussed in *Gulino II*, *see* 460 F.3d at 374–78, is reasonable. The interference test may be inappropriate in many settings, but where the SED, as the only party with any responsibility for creating and instituting two discriminatory exams, is absolved of liability at the expense of the BOE, who was forced by the SED to adopt the exam, the interference test may be the best way of ensuring an equitable outcome to this case. *Cf.* Michele A. Yankson, Note, *Barriers Operating in the Present: A Way to Rethink the Licensing Exception for Teacher Credentialing Examinations*, 89 N.Y.U. L. Rev. 1902, 1933 (2014) (arguing that the “narrow control-test analysis used by some courts” is insufficient to effectuate the remedial-based purpose of Title VII).

In addition, recent changes that greatly increase the level of the State's control over teachers may warrant consideration in future cases. Since the Second Circuit decided *Gulino II*, the SED has established much greater regulatory control over New York's public school teachers pursuant to the regulatory mandates associated with two federal education programs: No Child Left Behind and Race to the Top. *See* Yankson, *supra* at 1925–28 (2014) (discussing the ways in which No Child Left Behind gave states greater regulatory control over their teachers' job performance); Regina Umpstead, et al., *An Analysis of State Teacher Evaluation Laws Enacted in Response to the Federal Race to the Top Initiative*, 286 Ed. Law Rep. 795, 795 (2013) (noting how Race to the Top “dramatically changed the public school teacher employment landscape in a short timeframe”); *id.* at 803–13 (describing the teacher evaluation systems state agencies have implemented pursuant to Race to the Top); Robert S. Eitel et. al., *The Road to A National Curriculum: The Legal Aspects of the Common Core Standards, Race to the Top, and Conditional Waivers*, 13 Engage: J. Federalist Soc'y Prac. Groups 17, 21 (March 2012) (asserting that the Race to the Top persuaded most states to adopt the Common Core Standards, and discussing how the Common Core standards affect the way teachers teach). This level of control could possibly suffice to find that the SED is an employer of public school teachers under Title VII, based on the legal requirements set forth in *Gulino II*. *See* Yankson, *supra*, at 1928–32. *But see O'Connor v. Davis*, 126 F.3d 112, 115–16 (2d Cir. 1997) (holding that a “prerequisite” of being an employee pursuant to Title VII is that “the individual have been hired in the first instance,” and noting that a lack of remuneration by the entity in question can prove dispositive); *Gulino II*, 460 F.3d at 379 (stating that “plaintiffs fail[ed] to meet the threshold showing that SED hired and compensated them”).