

2003 WL 25764041

Only the Westlaw citation is currently available.
United States District Court,
S.D. New York.

Elsa GULINO, Mayling Ralph, Peter Wilds, and
Nla Greene, on behalf of themselves and all others
similarly situated, Plaintiffs,

v.

The BOARD OF EDUCATION OF the CITY
SCHOOL DISTRICT OF THE CITY OF NEW
YORK; and New York State Education
Department, Defendants.

No. 96 Civ. 8414(CBM). | Sept. 4, 2003.

Attorneys and Law Firms

Barbara J. Olshansky, Joshua Sohn, Piper Rudnick LLP,
Stephen G. Seliger, Seliger & Elkin, LTD. Joel M.
Hellman, for Plaintiff Class.

Michael A. Cardozo, Corporation Counsel of the City of
New York, by: Donald Sullivan, of Counsel, for
Defendant City Board of Education.

Eliot Spitzer, Attorney General of the State of New York,
by: Bruce B. McHale, Jane A. Conrad, Robert M. Blum,
William A. Jaffe, Assistant Attorneys General, of
Counsel, for Defendant State Education Department.

Opinion

FINDINGS OF FACT AND CONCLUSIONS OF LAW

MOTLEY, District Judge.

*1 The plaintiff class in this case is comprised of African American and Latino educators in the New York City public school system (“plaintiffs”) who have either lost their teaching licenses or have been prevented from obtaining a full teaching license because of the requirement that they achieve a satisfactory score on one of the challenged tests, the National Teacher Core Battery Exam (the “Core Battery”) or the Liberal Arts and Sciences Test of the New York State Teacher Certification Examination (“LAST”), the successor to the Core Battery. The tests were developed and administered by defendant New York State Education Department

(“SED”).

Plaintiffs alleged that the use of the tests as a requirement for obtaining permanent teaching certificates has an impermissible disparate impact on African–American and Latino teachers in New York City public schools, in violation of Title VII of the Civil Rights Act of 1964, § 701 *et seq.*, as amended, 42 U.S.C. § 2000e *et seq.* Plaintiffs also alleged that white test-takers passed both tests at a rate that is statistically significantly higher than the rates for African Americans and Latinos. Plaintiffs further alleged that the tests in question were misused and do not measure whether the test-takers are qualified to be teachers. Many of the plaintiffs have remained employed as teachers with defendant New York City Board of Education (“BOE” or “the Board”), some in the same classrooms teaching the same subjects. They have asserted, however, that they have suffered harm in that they have had their salaries reduced and have been denied employment opportunities, rights and benefits.

The parties to this action presented their evidence during an epic bench trial that lasted more than eight weeks and filled over 3,600 pages of trial transcript. This decision constitutes the findings of fact and conclusions of law of the court.

For the reasons that follow, although the court finds that defendants have not demonstrated that the LAST was properly validated according to professional guidelines and standards, the court does find that both the Core Battery and the LAST are job-related. Accordingly, the court finds for defendants.

INTRODUCTION

Disparate impact suits follow a three stage paradigm, as explained in multiple Supreme Court cases: (1) plaintiff must prove that the challenged practice has a disparate impact on a protected group; (2) once that showing is made, the burden shifts to defendant to demonstrate that the challenged practice bears a “manifest relationship to the employment in question”; and (3) if the defendant meets its burden, it prevails, unless plaintiff can offer an equally practicable, cost effective alternative practice which serves the same purpose as the challenged practice but does not result in the same disparate impact. *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975) (quoting *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973)); *see also Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971). Accordingly, these findings will

follow that three stage structure.

FINDINGS OF FACT

Background: Teacher Certification and Testing in New York

*2 1. Prior to January 1, 1991, teachers and prospective teachers in New York City were certified by the Board of Examiners of New York City, a City agency distinct from the Board. Tr. at 1668–69.

2. The license issued by the Board of Examiners for service in the New York City Board of Education schools on its face indicates that it is a conditional one, in that it is “subject to meeting minimum requirements” by a date certain and “subject to meeting maximum requirements” by a date certain, which is five years after the minimum requirements date. PL Ex. 1 at 5 (plaintiff Mayling Ralph–White’s City license). *See also* Tr. at 238. On its face, the license also states that it “may be used for service as a substitute until such time as a teacher receives a regular appointment.” Pl.Ex. 1 at 5.

3. Provided minimum conditions were met, there was no expiration date for the license. *See* Pl.Ex. 1 at 5. Because there was no expiration date and because the license stated that it could be used to perform substitute service, a person working under that license could teach full time as a substitute. Tr. at 868–69. Once maximum conditions for licensure were met, the license could be used to obtain a position as a regularly appointed teacher in New York City. Pl.Ex. 1 at 4.

4. To obtain this conditional license, teacher candidates were required to pass either an “open examination” or, at certain periods of time, a “closed examination.” Tr. at 195. The “closed examination” was created by State law in 1986, and the examination consisted only of an oral interview. *See* Tr. at 200–01, 234. This closed examination legislation was passed in response to a shortage of teachers in New York City. Tr. at 918. The “open examination” was available to all teacher candidates who had or were in the process of attaining a college degree; the test consisted of a content knowledge short answer section, a written English portion and an interview portion during which the candidate was asked questions pertaining to pedagogy. Tr. at 1653–58, 1662–63.

5. Only certain people were eligible to take the “closed examination”; the test was given only to those who had

worked in the New York City public schools for two years as substitute teachers under a “temporary per diem certificate.” Tr. at 200–01, 1656.

6. The temporary per diem certificate was created because of the severe teacher shortage and the need to fill classrooms with certified teachers. Tr. at 915. To obtain a temporary per diem certificate, teacher candidates had to write a 300 word essay and pass an oral examination. Tr. at 914. The holder of a temporary per diem certificate was eligible to teach for one year, and the license could be renewed again in subsequent years. *See* PL Ex. 1 at 1.

7. The closed examination was basically an interview test; Marie DeCanio, Senior Deputy Executive Director for Human Resources at the Board during the relevant time periods, testified that “it was really a test that was to get people who were working as sub[stitute]s to have an easier method into the system as the initial certificate, and then they would have to complete their five-year requirements after that.” Tr. at 239.

*3 8. If a teacher candidate met the requirements to take the closed examination and passed that examination, he or she would get a conditional City license, which would be no different than the conditional license granted to a person who had passed the open examination; each of these licenses would be subject to meeting minimum conditions by a date certain, and maximum conditions by five years later. *See* Tr. at 235–38; PL Ex. 1 at 5.

9. Teachers and teacher candidates who were going to receive a conditional City license as a result of a closed examination were notified in writing that the license that would “be issued subject to: ... [m]eeting the eligibility requirements in full by” a date certain, which was five years after the notification that the license would be issued. Pl.Ex. 1 at 4.

10. There was one crucial distinction between persons who received their conditional City license via the open examination and persons who received that license via the closed examination: by State law, if a person passed the open examination and took more than five years to meet his or her maximum requirements, the City license could be restored; but the City license of a person who passed the closed exam could not be restored if he or she took more than five years to meet his or her maximum requirements. Tr. at 956. In fact, the statute that established the closed examination declared that candidates who achieved their license via that route could not have the license reinstated if they failed to achieve maximum requirements in a timely fashion. Board Ex. L at 2.

11. Although the maximum licensure requirements

included the tests at issue in this litigation, there were other maximum requirements that also had to be met in the same five-year period, including, for example, obtaining a Master's degree. Tr. at 199.

September 1, 1984: SED Imposes the Core Battery as a Requirement

12. City licensing standards have always been subject to State review and the City has always been required to meet State minimum licensing standards. Marie DeCanio testified that prior to January 1, 1991, "we were subject to state requirements that we had to at least meet [the State's] minimum" requirements, Tr. at 243–44, and that even when the "actual [State] certificate was not required[,] ... comparable background and experience was required" of City Board teachers by SED. Tr. at 193. The State Education law additionally "required that the City Board ... adopt regulations for the licensure of teachers in the city that [were] substantially equivalent or not less than those required by the state." Tr. at 1725. Whatever licensing requirements the Board adopted had to be approved by SED. Tr. at 1005; Board Ex. N. Thus, under the statutory and regulatory scheme, SED "ha[d] the responsibility of establishing the qualifications for ... the permanent certificate as well as issuing the certificate." Tr. at 1006.

13. State certification can be obtained on a provisional basis, meaning that the certification is good for five years, and on a permanent basis, meaning that the certification is valid for life. Tr. at 1710.

*4 14. In September 1984, a SED regulation stated that: "All applicants for a provisional certificate must submit evidence of having achieved a level of performance satisfactory to the [State Education] Commissioner on examinations acceptable to the Commissioner." SED Ex. 17 at SED 145; Board Ex. N. The regulation declared:

A license issued ... by the superintendent of schools of the city school district of the City of Buffalo or of the City of New York shall be valid for service in any school district in the State for a position for which a comparable State certificate is required, provided such license has been issued upon qualifications which are at least substantially equivalent to those required for [State certification] ... pursuant to these

regulations. Comparability of the State certificate pursuant to these regulations and the certificate granted by such superintendents shall be determined by the [C]ommissioner [of SED].

Board Ex. N, ¶ (j). This regulation also applied to persons who were seeking permanent State certification. SED Ex. 17 at SED 146.

15. There was a limited exception to this regulation, for teachers in New York City and Buffalo, in that teachers in those systems who were already "regularly licensed by the boards of education in New York City and the City of Buffalo as of September 1, 1984, [were] ... not [required] to take the Core Battery Tests as long as the respective license [was] valid at the time of application for a State certificate." *Id.*

16. The Commissioner of Education is the chief executive officer of SED and is responsible for implementing the Regents' policy directives. Ed. Law § 305; Tr. at 3298. In its 1977 report, the Commissioner's task force on teaching recommended that teachers be required to complete a broad course of study in the liberal arts and pass an initial licensing test in the liberal arts, noting that such a test would not be intended to predict teacher performance, but rather would test minimum competencies. SED Ex. 1 at SED 23; Tr. at 1898. Hearings across the state confirmed that better qualified teachers were needed to improve student performance. Tr. at 1731–32. This recommendation was consistent with a national movement emphasizing the value of a general education and knowledge of basic liberal arts for teachers. SED Ex. 800 at 14–15.

17. On April 24, 1981, the Regents instituted a liberal arts and sciences testing requirement for candidates seeking State certification on and after September 2, 1984. 8 NYCRR, § 80.2(q), Tr. at 1732. Because prior to January 1, 1991, New York City teachers were licensed exclusively by the City Board of Examiners and were not subject to State licensing requirements, this regulation did not apply to New York City. Ed. Law § 2590–j(2); Chapter 650 of the Laws of 1990; Tr. at 1724–25, 1736. The regulation delegated to the Commissioner the responsibility for selecting appropriate tests and setting the passing score. SED Ex. 11. SED contracted with Educational Testing Service (ETS) to assess whether the Core Battery would be valid for certifying New York teachers. Tr. at 2209–10, 2273–75; SED Exs. 10, 12.

*5 18. In 1984, SED adopted the Core Battery as the test of the liberal arts and sciences for candidates for teacher

certification. Tr. at 1732; SED Ex. 17 at SED 145. Prior to making its selection of the Core Battery, SED “reviewed several alternatives” and concluded that the Core Battery was “the only existing teacher-related test.” PL.Ex. 24 at GUL 01080. SED made this selection after a validity study had been conducted and the Core Battery “had been found valid for use in New York State as a certification examination.” SED Ex. 17 at SED 145.

19. Once SED made this selection, administrators at the Board concluded that they were required to have their candidates for licensure also pass the Core Battery. Tr. at 232, 943, 1236, 1256, 1699. Dr. Charles C. Mackey, Jr., who held several different positions high up in SED’s Office of Teaching during the relevant times, testified that SED’s decision to use the Core Battery as a screening device was in effect imposed on the Board, even prior to the 1991 law that eliminated the Board of Examiners as the City’s licensing arm:

Q: But just to clarify, the state is responsible for setting the requirements, is that right, sir?

A: Yes.

Q: And that would apply even to people that are hired prior to 1991, is that right, sir?

A: Well, in that the—in that the requirements for a regular New York City license would be akin to or the equivalent of what the state required at the time for a permanent certificate, the answer is yes.

Tr. at 1006–07.

20. Dr. Mackey further testified that the Core Battery and the LAST were the most appropriate tests to use to measure a candidate’s knowledge of the liberal arts and sciences for purposes of achieving certification in New York State. Tr. at 2349. In other words, there was no “substantial equivalent” to the Core Battery prior to the creation of the LAST. SED told the Board that it was required to use the Core Battery. Tr. at 3483–84.

21. Following the State’s imposition of the Core Battery as a minimum requirement for State certification, and therefore a minimum requirement for City licensure (except for those teachers who had already obtained a regular City license), in 1985 the City requested, by letter, permission from SED to have the Core Battery made a maximum licensure requirement. PL Ex. 30. The request asked SED to

permit [the Board] to confer a license upon a candidate who fulfills all academic and

professional requirements other than the Core Battery accomplishment, with the understanding that the license will be terminated upon failure to achieve whatever else is prescribed (including the Core Battery) within the five-year interval provided.

PL Ex. 30.

22. The City Board made this request because it did not want to require teacher candidates to take and pass multiple tests in a compressed time frame and also because the Board was facing difficulty in staffing teaching positions. Tr. at 254; PL.Ex. 30. The Board also noted that it was sometimes forced to staff positions through the temporary per diem certificate, which was “clearly a less desirable vehicle for bringing in new personnel.” PL Ex. 30.

*6 23. SED granted the Board’s request for a deferral of the Core Battery as a minimum requirement. PL.Ex. 31. The agreement memorialized an understanding that the Core Battery would “be a minimum requirement with delayed implementation to allow [City] teachers to have a benefit of taking the test a number of times ... [while] they could continue working.” Tr. at 213.

24 In an internal 1989 memorandum, SED recited its understanding that the decision to require passage of the Core Battery by City teachers was a SED requirement, that the City had to ask the State’s permission for a deferral of this requirement and that City teachers who failed to pass the examination in a timely manner would be precluded from continued employment as a regular teacher. PL Ex. 32; see Tr. 973 (“It is a memorandum ... indicating ... the nature of the approval that had been or authorization that had been granted some four years earlier to the city to defer the requirement of the Core Battery from the entry level to other requirements for regular licensure in the city”).

Teacher Assignments in New York City: Before January 1, 1991

25. Persons who received their initial New York City license by taking the closed examination, because they were by definition substitute teachers already in the system, were typically placed in a vacancy in the school where they were serving as substitutes. Tr. at 1667. If there were no vacancy in that school, the district superintendent would place the person in another vacancy in the district. *Id.*

26. Persons who achieved their initial New York City license by passing the open examination were hired through a rank-ordered system, which matched job openings with the persons on the eligible for hire lists who received the highest test scores. Tr. at 1663. Eligibility lists were maintained for each test administration, and an eligibility list stayed open for seven years. Tr. at 1663–65. The City exhausted all available candidates on earlier lists before moving to later lists, but, within each list, the persons with the highest scores were considered first for open positions. Tr. at 1663–64.

27. A candidate who declined an appointment was required to document the asserted good cause for their unavailability and could not decline an appointment merely because he or she did not want to teach in a particular school. Tr. at 1681. The declining teacher was required to establish that he or she met one of a limited number of acceptable reasons for declining: travel distance to assignment; caring for a child or sick family member; pregnancy; being under contract to a different school system; or having moved out of New York State. Tr. at 1681–83.

28. Teachers who declined positions without good cause were removed from the eligible list and lost the right to appointment, meaning that they would no longer be offered any position. Tr. at 1682–83. Such teachers, however, did not lose their New York City licenses, because they had a property right to those licenses. Tr. at 1683. Similarly, teachers who failed to achieve maximum licensure requirements would not lose their substitute teacher’s license once they had met the minimum licensure requirements; they would, however, lose the chance to receive a regular appointment. *See, e.g.*, Pl.Ex. 1 (fourth page).

*7 29. It has always been the case, however, that, pursuant to State law, the City could not employ as regularly appointed teachers persons who did not meet the State requirements for teachers. Tr. at 214.

Changes Effective January 1, 1991: Elimination of Board of Examiners; Elimination of Closed Examination; and State Certification Becomes a Minimum Requirement

30. Effective January 1, 1991, by State law, the Board of Examiners was eliminated as a licensing entity and its licensing responsibilities were transferred to the Board. SED Ex. 58(c) at SED 1898–1901, 1892. That law, Chapter 650 of the Laws of 1990, dictated that City licensure examinations would be an interview examination only. Tr. at 1669; SED Ex. 58(c). Under this

same law, effective January 1, 1991, all teachers in New York State were required to obtain State certification. Tr. at 925. Thus, after that date, New York City teachers were required to have State certification in order to obtain a New York City license. See Tr. at 213–14, 242. This change, however, did not affect teachers who already had regular New York City licenses. Tr. at 214–15.

31. These teachers were not subject to State certification requirements—except as those requirements related to the individual teacher’s efforts to attain maximum requirements for regular City licensure. See Tr. at 926; see also SED Ex. 17 at SED 146; Pl.Ex. 30. Accordingly, a person who had a license to teach as a substitute in New York City would continue to hold that license and be able to teach under that license, even though he or she lacked State certification. See Tr. at 214–15. Similarly, a person who was already a regularly appointed teacher and who possessed a full (as opposed to conditional) New York City license did not have to obtain State certification and could maintain his or her appointment. SED Ex. 58(c). Teachers who held only a conditional New York City license, however, would be required, under the State regulation, to obtain State certification. See SED Ex. 17 at SED 146.

32. Further, as a result of this law, the Core Battery, in effect, became a minimum requirement for State certification. Tr. at 213–14. Once passage of the Core Battery became, again, a minimum requirement for State certification, it became a minimum requirement for City licensure, pursuant to SED regulations. Tr. at 213–14.

Creation of ORPAL

33. In late 1990, in anticipation of the January 1991 legislative change, the Board created an office, known as the Office of Recruitment, Personnel, Assessment and Licensing (“ORPAL”), which was responsible for, among other items, monitoring the status of teachers’ efforts to achieve maximum requirements for licensure. Tr. at 980, 1997–70; see Pl.Ex. 273. The ORPAL office implemented a change in record-keeping and converted manual records into computerized ones, in order to keep track of a teacher’s progress toward meeting maximum requirements. Tr. at 870.

*8 34. Candidates who first applied for a City license after January 1, 1991 and who possessed State provisional certification were placed on a qualifying-eligible list, which was an alphabetical listing (as opposed to the rank order list that had been created under the Board of Examiners tests). Tr. at 1670. As was the case with the Board of Examiners’ lists, teacher candidates had a right to stay on this list for seven years. Tr. at 1669. Names

were drawn at random from the post-January 1, 1991 list to fill vacancies. Tr. at 1669–70.

35. Based on “grandperson” provisions in the January 1, 1991 effective legislation, people who applied for City licenses before that date were not subject to the new assignment procedures and received preferential treatment over the new applicants who had passed the Core Battery certification examination and received State provisional certification. Tr. at 1671–72. Persons who applied pre-1991 were maintained on the rank order lists and, when a position opened up, ORPAL went through all the rank order lists in an attempt to fill that vacancy, before it considered people who were on the new pool list. *Id.*

36. As with the pre-1991 rules, a person could not decline an appointment to a school because he or she did not want to work there. Tr. at 1690. If a person declined an appointment to a particular school, his or her name would be removed from the list of prospective teachers in the system and could only be restored to that list if he or she agreed to go to a comparably difficult-to-staff school. Tr. at 1690–91.

State Temporary Licenses

37. After the 1991 law, a teacher candidate in New York State, including New York City, could still teach without having passed the requirements for State certification, including achieving a passing score on the Core Battery, provided he or she had a State temporary license. Tr. at 925–26. A State temporary license could be renewed up to three years, although this period was extended in certain instances. Tr. at 889. Temporary licenses have been renewed repeatedly because of the severe teacher shortage in New York City. Tr. at 929.

38. Because State certification had become, by law, a requirement for all teachers as of January 1, 1991, after that date the State took “a more visible active role in monitoring the employment status, the licensure status of teachers in the city.” Tr. at 979, 1006–07.

39. In 1991, the City began applying for temporary State licenses on behalf of many of the City’s teachers, so that they could teach in the City system. Tr. at 926. The City found it necessary to apply for these temporary licenses in large numbers. *Id.* The severe shortage of qualified teachers in New York City has led to the City seeking renewals on temporary licenses, because the alternative would be that classrooms would go unstaffed. Tr. at 247–48.

40. In order to obtain these temporary licenses, the Board was required by SED to certify that: (1) there were no

certified teachers available to teach in the particular school or district where the temporary license was to be used; and (2) the teacher proposed by the Board was working toward State certification. Tr. at 928. The individual teacher was required to sign the temporary license application and to indicate that he or she was working to achieve full State certification. Tr. at 2324, 929; SED Ex. 812.

Enforcement of the Core Battery/LAST Requirement

*9 41. During the years leading up to 1991, SED repeatedly contacted the Board and reminded it to make sure that teachers had satisfied requirements for full, regular licensure. Tr. at 979–80. This effort on the part of SED became more pronounced as the City began to apply for temporary licenses in greater numbers. *Id.* Simultaneously, ORPAL’s efforts to mechanize its records, and therefore enable greater accountability, began to show results. *See* Pl.Ex. 273.

42. Teachers were often given additional time, beyond the five-year period announced on their conditional license, to achieve maximum requirements, primarily as a result of ORPAL’s efforts to mechanize its records. *Compare* Pl.Ex. 1 (fifth page) (named plaintiff’s license expired in May 1994 by its own terms), *with* Pl.Ex. 273 (same named plaintiff given until June 1995 to show that she had met maximum requirements).

43. In 1993, SED began to phase out the use of the Core Battery and phase in the use of the LAST. Tr. at 2316. Teachers could take either test to satisfy the requirement. Tr. at 660. SED decided which teacher titles were eligible to take either of these examinations. Tr. at 2351. As a result of SED’s decision, many teachers could pass either the Core Battery or the LAST to satisfy their maximum requirements for certification. Tr. at 2316.

44. SED directed the Board to take action with respect to teachers who failed to timely achieve maximum licensure requirements. Tr. at 2350.

PART I: PLAINTIFFS’ PRIMA FACIE CASE

45. Plaintiffs and SED each retained a statistical expert who testified with respect to the differences in pass rates on each of the tests. Dr. Thomas A. DiPrete, testified on behalf of plaintiffs. Tr. at 462–576. Dr. Joan Haworth testified on behalf of SED. Tr. at 2465–2566.

The LAST'S Disparate Effect

46. Among people employed in the City's public schools, the difference in pass rates between White and African-American first-time test-takers of the LAST was statistically significant at more than fifty-nine standard deviations, Tr. at 480–81; Pl. Exs. 222, ¶ 22 at 12; Pl.Ex. 222, App. I, Table A12 at 22–23, and the difference in pass rates between White and Latino first-time test-takers was statistically significant at more than sixty-five standard deviations. Tr. at 481; Pl. Exs. 222, ¶ 22 at 12; Pl.Ex. 222, App. I, Table A12 at 22–23.

47. Among people employed in the City's public schools, the difference in pass rates between White and African-American test-takers—looking at whether test-takers ever passed the LAST—was statistically significant at more than eighteen standard deviations, Tr. at 499–500; PL Ex. 227, App. Table 14 at 46, and the difference in pass rates between White and Latino test-takers was statistically significant at more than nineteen standard deviations. Tr. at 500; Pl.Ex. 227, App. Table 14 at 46.

48. Among test-takers across the State, the difference in pass rates between White and African-American first-time test-takers of the LAST was statistically significant at more than seventy-five standard deviations, Pl.Ex. 222, ¶ 20, at 11–12; Pl.Ex. 222, App. I, Table A10 at 22, and the difference in pass rates between White and Latino first-time test-takers was statistically significant at more than seventy-five standard deviations. Pl.Ex. 222, ¶ 20, at 11–12; Pl.Ex. 222, App. I, Table A10 at 22.

*10 49. Among test-takers across the State, the difference in pass rates between White and African-American test-takers—looking at whether test-takers ever passed the LAST—was statistically significant at more than forty-two standard deviations, PL Ex. 227, App. Table 12 at 44, and the difference in pass rates between White and Latino test-takers was statistically significant at more than forty-three standard deviations. PL Ex. 227, App. Table 12 at 44.

The Core Battery's Disparate Effects

50. Among people employed in the City's public schools, the difference in pass rates between White and African-American first-time test-takers of the Core Battery was statistically significant at more than eleven standard deviations, Tr. at 478–79; Pl. Exs. 222, ¶ 15 at 9–10; Pl.Ex.222, App. I, Table A7 at 20–21, and the difference in pass rates between White and Latino first-time test-takers was statistically significant at more than twelve standard deviations. Tr. at 479; Pl.Ex. 222, ¶ 15 at 9–10; Pl.Ex. 222 App. I, Table A7 at 20–21.

51. Among people employed in the City's public schools, the difference in pass rates between White and African-American test-takers—looking at whether test-takers ever passed the Core Battery—was statistically significant at more than thirty-six standard deviations, Tr. at 495–97, Pl.Ex. 227 at 34, and the difference in pass rates between White and Latino test-takers was statistically significant at thirty-seven standard deviations. Tr. at 497–98; PL Ex. 227 at 34.

52. Among test-takers across the State, the difference in pass rates between White and African-American first-time test-takers of the Core Battery was statistically significant at more than thirty standard deviations, Pl.Ex. 222, ¶ 12 at 8; Pl.Ex. 222, App. I, Table A5 at 18, and the difference in pass rates for White and Latino first-time test-takers was statistically significant at more than thirty standard deviations. Pl.Ex. 222, ¶ 12 at 8; Pl.Ex. 222, App. I, Table A5 at 18.

The Four-Fifth's Rule

53. Section 4(D) of the Equal Employment Opportunity Commission's Uniform Guidelines (the "Uniform Guidelines") provides that a selection rate for a protected group that is greater than four-fifths of the rate for the highest group will generally not be regarded by federal enforcement agencies as evidence of adverse impact. However, such differences may constitute adverse impact where they are significant in both statistical and practical terms. 29 C.F.R. Section 1607.4(D). This guideline is referred to as the "Four-Fifths Rule." Dr. DiPrete testified that the "Four-Fifths Rule" is not a measure of disparity that is commonly used in the scientific or academic community. Tr. at 488.

54. The pass rates on the Core Battery for African-American and Latino test-takers are less than four-fifths of the pass rates for White test-takers with regard to all of the groups Dr. DiPrete considered in his report. Tr. at 491–492, 2543–45, 2544; Pl.Ex. 222, ¶ 25 at 13.

*11 55. The pass rates on the LAST for African-American and Latino first-time test-takers are less than four-fifths of the pass rate for White test-takers. Tr. at 491–492, 2545–2546, 2553; PL Ex. 222 ¶ 25 at 13. Moreover, with respect to eventual pass rates on the LAST, the disparities in those pass rates have both statistical significance (at more than eighteen standard deviations as noted above), and practical significance for those populations. Dr. Haworth testified that if the eventual pass rates on the LAST had been as high for African-American and Latino teachers in City schools as it has been for White teachers, more than one thousand additional African-American teachers and more than

several hundred additional Latino teachers would have passed the LAST. Tr. at 2550–2553.

56. The differences in pass rates on the Core Battery between White test-takers and African–American and Latino test-takers, respectively, also have practical significance. Dr. Haworth testified that if African–American test-takers had the same pass rate as White test-takers, approximately 2,000 additional African–American test-takers would have passed the Core Battery; if Latino test-takers had the same pass rate as White test-takers, approximately 2,000 additional Latino test-takers would have passed the Core Battery. Tr. at 2546–2550.

57. Accordingly, this Court finds that under the Four–Fifths Rule, the LAST and the Core Battery have an adverse impact on African–American and Latino teachers.

Defendants’ Rebuttal

58. Dr. DiPrete testified that it was appropriate to analyze the pass rates of first-time test-takers because there were adverse consequences to not passing the tests, even on the first try. Tr. at 485–86. He concluded that, since passage of one of the tests was a necessary step in becoming certified, a teacher who failed the test was clearly worse off than one who passed and therefore suffered adverse consequences from failing the test for this reason alone. Tr. at 485–86, 520–21. By contrast, Dr. Joan Haworth, SED’s expert, testified that in determining whether the tests have a disparate impact, she focused on whether test-takers ever passed the tests. Tr. at 2486. She based her preference on her assumption that there were no adverse consequences to failing the Core Battery or LAST on the first or subsequent tries, so long as the individual eventually passed one of the tests. Tr. at 2484–2486; 2540–41. Dr. Haworth’s “ever pass” analysis, however, is based on erroneous assumptions and is neither reliable nor credible for the reasons explained in the following paragraphs.

59. There are a number of adverse consequences to each failure of the tests. A person who failed either of the tests had to pay a fee each time he or she retook the tests, as well as an annual fee for issuance or renewal of a State Temporary License. Tr. at 2354. Although Dr. Haworth admitted that the payment of additional fees constitutes an adverse consequence of failing the tests, she did not consider this fact in her analysis. Tr. at 2536–37; 2541.

*12 60. In addition, some class members who failed either of the tests had their licenses revoked and were demoted to substitute status, Tr. at 963–65; 1221–22; Pl.Ex. 273, with an attendant decrease in salary of between \$2000 and \$20,000 per year. Tr. at 906–907.

61. After 1990, other class members had to pass either the Core Battery or LAST in order to receive provisional state certification and to be eligible for a regular license from BOE. Tr. at 884–85. Without a City license, a teacher is ineligible for a permanent appointment and tenure and therefore cannot advance up the salary scale and accrue seniority in the City School District. Tr. at 885–887. In addition, teachers without City licenses are required to apply for an annual State Temporary License in order to continue teaching and are eligible to teach only if they are nominated to fill a vacancy in a shortage area for which certified teachers are not available. Tr. at 887, 245, 2353–54.

62. There are also significant differences between the employment rights and benefits of certified teachers and those of substitute teachers. A certified teacher has a greater ability to advance on the salary scale than a substitute teacher, Tr. at 865, has priority for placement ahead of a substitute in the City School District, Tr. at 865, has greater system-wide seniority, Tr. at 866, and has the right to secure another full-time position in the district if her position is eliminated. Tr. at 865–66; SED Ex. 237, Art. 17, ¶ B at 105. If a school is in an excessing condition, a certified teacher can bump any substitute teacher and take her position. Tr. at 866; SED Ex. 237, Art.5, ¶ C at 15–16, Art. 17, ¶ A(3) at 104. A substitute teacher’s salary cannot exceed salary step 4A. Tr. at 866. Although Dr. Haworth admitted that the loss of such rights is an adverse consequence from failing the tests, she did not take into account the loss of these rights in her analysis. Tr. at 2533, 2536.

63. Moreover, while Dr. Haworth acknowledged that an enhanced risk of losing one’s job would be an adverse consequence of failing the Core Battery or LAST, she did not take this fact into account in her analysis. Tr. at 2535–36. Dr. Haworth did not consider the psychological injury suffered by teachers when they failed either of the tests even the first time. Tr. at 629, 2537.

64. Dr. Haworth’s analysis is inconsistent with her own factual assumptions. Dr. Haworth assumed that teachers had a five-year period during which to pass either of the tests, and that there were no adverse consequences so long as the teacher passed within this period. Tr. at 2483–85. However, Dr. Haworth never did an analysis based on this five-year period. Tr. at 2539–41. Instead, she conducted her analysis as though a test-taker would have suffered no adverse consequences if he or she passed either of the tests at any time during the sixteen-year period from 1984 (for the Core Battery) and from 1993 (for the LAST) up to the year 2000. Tr. at 2538–39, 2546.

*13 65. Given that the evidence shows both statistical and

practical significance, this court finds that Dr. Haworth's testimony does not rebut the evidence of disparate impact with respect to the LAST and the Core Battery. Accordingly, the court finds that defendants' use of the tests had a significant adverse impact on African American and Latino teachers in New York City.

PART II: JOB RELATEDNESS

66. Education measurement professionals consider whether an inference drawn from a test is valid by examining the evidence that the test measures what it is designed to measure. SED Ex. 800 at 27–32; Tr. at 3012–18, 425–26, 1123. In the instant case, the inference being drawn is whether a teacher possesses the knowledge and skills deemed necessary by the teaching profession and the Regents to be provisionally certified.¹ SED Ex. 800 at 29; Tr. at 3012–13. Experts' validity judgments are made based on an accumulation of evidence regarding how the test was developed, as well as post-administration statistics indicating its reliability. Tr. at 3011–20; SED Ex. 800 at 29–30.

The Core Battery Tests

67. The content evidence of the validity of New York's certification decisions based on candidates' passage or failure of the Core Battery tests, as well as the professional opinion of Dr. William Mehrens, SED's expert witness on testing and test validation, demonstrate that the State's certification decisions based on that test was both valid and reliable.

68. ETS's teacher certification testing program was overseen by the National Teacher Exam Policy Council (the "NTE Policy Council"), a group of 12 teachers and education faculty members from across the Country, including one New York educator. Tr. at 1919–20, 3021; SED Ex. 10 at GUL 129.

69. In the late 1970s, the NTE Policy Council decided to update its Common Examinations. The revised test series, the Core Battery, was developed from 1978 to 1982, and consisted of three separate two-hour tests in Communication Skills (including reading, writing and listening), General Knowledge (social studies, science, math, literature and fine arts), and Professional Knowledge (pedagogy). Tr. at 1913, 1917–20, 1933; SED 29 at E 418. The areas tested on the General Knowledge test were the same as those previously tested on the Common Examinations, plus a listening skills test and an essay test. Tr. at 1932–33.

70. Dr. Carolyn Massad, the ETS employee responsible for the initial development of the Core Battery tests, has a Ph.D. in educational psychology and is familiar with the AERA/APA/ NCME standards. Tr. at 1905–09; SED Ex. 839.

71. As part of the initial development, the NTE Policy Council commissioned papers regarding the knowledge and skills necessary for all teachers. Tr. at 1925–27. ETS collected a wide range of materials regarding teacher preparation and the teaching profession, including college course catalogues and articles from professional associations. Tr. at 1925, 1928, 1973; SED Ex. 29 at 422–425. ETS also collected materials for each of the content areas to be assessed and pedagogical resources regarding the methods of teaching each of the covered subjects. Tr. at 1929.

*14 72. The NTE Policy Council convened advisory committees of experienced educators for each of the three tests being developed. Committees were recruited nationwide through deans of education colleges, state licensing agencies and professional education associations. Tr. at 1924, 1930–31. The General Knowledge test advisory committee consisted of ten content experts in science education, mathematics education, social studies education, and literature and fine arts education. Tr. at 1930. The advisory committees for the Communication Skills and Professional Knowledge tests each consisted of eight members with expertise in reading, writing and listening skills and pedagogical experience, respectively. *Id.* The advisory committees reviewed the literature and curriculum surveys and then developed "test specifications", a blueprint of the test content, which were then mailed out in survey form. Tr. at 1924, 1930–31, 1934, 3025; see test specifications in SED Ex. 20, App. E. Through this survey, approximately 3,000 educators nationwide provided input on the proposed Core Battery tests' content. SED Ex. 29 at E 418; Tr. at 1934–38, 2105, 3022–25.

73. In the course of the Core Battery tests' development, determinations were made regarding the level of difficulty of the tests. Tr. at 1927. For example, the level at which mathematics knowledge on the Core Battery General Knowledge test was measured was comparable to what would have been acquired by the completion of high school or possibly one year of college. SED Ex. 36 at 1754; Tr. at 1953. Statisticians and psychometricians assisted in determining the number of questions necessary to reliably assess an area of content knowledge or a skill. Tr. at 1927, 1933, 1938, 1940–41, 1974; SED Ex. 20, App. E–12–16; SED Ex. 21 at GUL 144.

74. One hundred thirty-three subject matter experts were

involved in drafting over 2,000 test questions (“items”) for the Core Battery tests. SED Ex. 29 at E 418; Tr. at 1941–43, 1979, 3029–30. Each item was subjected to multiple reviews, including a “sensitivity review” by individuals trained to detect potential bias. SED Ex. 2; Tr. at 1944–47.

75. Following these multiple reviews, test questions were assembled into “test forms,” which were subjected to further review. SED Ex. 2 at 5–7; Tr. at 1949–50, 1999. Each test form had a certain number of questions in common, called an “equating block”, enabling ETS to ensure that each test form was of comparable difficulty. Tr. at 1948; 2277, 3044. In addition, entire test forms were periodically re-administered. Tr. at 1980–81. Pretests of the Core Battery test questions were then administered in approximately 200 teacher colleges in 49 states. SED Ex. 29 at E 418; Tr. at 2106.

76. The Core Battery test development process was fully documented at the time.. Tr. at 1975–77, 1998–99. There was a fire at the storage facility of ETS’s storage contractor in the mid–1990s, however, and most of these documents did not survive. Tr. at 1923, 1969–70, 2301.

*15 77. The Core Battery was first administered for teacher certification purposes in November 1982. Tr. at 1933; SED Ex. 10 at GUL 129. A set of three representative Core Battery tests was received in evidence as SED Ex. 329.² The Core Battery tests were administered multiple times each year at various locations around the country. SED Exs. 17, 36; Tr. at 1950–51; SED Ex. 36 at E 1743. In 1988, twenty states used one or more of the Core Battery tests for teacher certification purposes. SED Ex. 36 at E 1767–1773.

78. The Core Battery tests were designed to determine whether teacher candidates possessed the academic knowledge and skills being assessed: “The NTE Programs tests are designed to measure academic preparation for teaching and not the act of teaching.” SED Ex. 29 at E 417; Tr. at 1941, 2000. The Registration Bulletin described the test content, the level of difficulty of the content, how to register for the test and which states accepted the Core Battery for credentialing purposes. SED Ex. 36; Tr. at 1952–54. ETS also published guides to each of the Core Battery tests. SED Ex. 36 at 1753–54; SED Ex. 29 at E 419.

79. Following a test administration, a candidate was sent a score report indicating his or her performance in each content area, such as English, mathematics, science, social studies or literature and fine arts. Tr. at 1955. ETS retained test scores for five years and a candidate could ask that scores be forwarded to a state or other credentialing agency after taking the test and being

advised of her score. SED Ex. 36 at E 1750.

80. Commencing in the mid–1980s, after each test administration, ETS performed a statistical analysis called Differential Item Functioning (“DIF”) to determine whether different racial or ethnic groups performed differently on individual test items. Tr. at 1958–62, 1978–79; SED Exs. 191–94, 196–97.

81. When registering for the test, candidates could elect to provide ETS with information such as where they had received their teacher preparation. SED Ex. 36 at E 1746. ETS used this information for statistical analyses of candidates’ performance and research regarding its testing programs, including analyses of the factors associated with different performance levels on the Core Battery tests. SED Exs. 186–197; Tr. at 1956–57.

82. In March 1983, New York contracted with ETS to perform a validity study to ascertain whether the Core Battery tests were appropriate for use as New York State’s initial teacher certification examination, Tr. at 2210, 2274–75; SED Exs. 10, 12, and to enlist the expertise of New York teachers and education college faculty members in calculating a proposed passing score for the consideration of the Commissioner of Education in the event that the Core Battery tests were adopted for use in New York. Tr. at 2209, 2213, 2268–69, 2273, 1966, 1968–69, 2190–91, 3038–56; SED Exs. 29 at E 418 and 36 at E 1750. This validity and standard setting study was conducted under the direction of Dr. Jane Faggen. Tr. at 2209–15. Dr. Faggen has a Ph.D. in educational psychology and is familiar with the joint APA/AERA/NCME standards. Tr. at 2203–05.

*16 83. The appropriateness of the use of the Core Battery for certification in New York was determined through the participation of New York public school teachers and administrators and college faculty from New York teacher preparation colleges on three panels: Content Review; Job Relevance; and Knowledge Estimation. Tr. at 2215–19; SED Ex. 20, App. F–1. There were also separate panels for the Communication Skills and Professional Knowledge tests and for the four different parts of the General Knowledge test. Tr. at 2219, 2309; SED Ex. 20 at 2. Dr. Vincent Gazzetta, SED’s Director of Teacher Certification, was SED’s liaison with Dr. Faggen on the validity study. Tr. at 2216–18.

84. Dr. Gazzetta oversaw the recruitment and selection of panel members. He made significant efforts to recruit participation from a diverse group of educators—racially, ethnically, geographically, in terms of school district type, grade level, experience level and subject matter. Tr. at 2216–17, 2272; SED Ex. 20, Appendices B–34, C and D; SED Exs. 6, 13. Three hundred forty-one New York

public school teachers, administrators and education college faculty members participated on the various panels, which met during the spring of 1983 at five different sites throughout the state. SED Exs. 13 and 20, App. D; Tr. at 2218–20.

85. Panel members were provided with a general orientation and were trained on their specific task. Tr. at 2237–38; SED Ex. 20, Appendices E–23, F, G. They were instructed to carefully analyze the knowledge and skills being assessed in each question to consider whether the knowledge and skill being assessed “is important for the beginning teacher.” SED Ex. 20 at E–23; Tr. at 2238–40; SED Ex. 10 at GUL 129.

86. The Content Review Panel, composed solely of teacher college faculty, assessed whether teacher candidates had the opportunity to acquire the knowledge and skills tested on the Core Battery tests in their teacher preparation programs. Tr. at 2219–29; SED Ex. 20 and App. E. This panel determined that individuals attending New York teacher education programs would have been exposed to nearly all of the knowledge and skills assessed on the Core Battery. SED Ex. 20 Executive Summary and p. V.8–10; Tr. at 2230–33, 3042, 3430–31.

87. The Job Relevance Panel determined the relevance of the content knowledge and skills assessed on all three Core Battery tests to the job of a New York public school teacher. SED Ex. 20 at SED 195; Tr. at 2233. The Job Relevance Panel consisted solely of teachers and administrators in New York public schools. Tr. at 2219. For each test question on two test forms, the panel made a judgment whether the knowledge or skill being measured was “crucial,” “important,” “questionable” or “not relevant” to the job of a beginning teacher in New York. Tr. at 2233–35; SED Ex. 20 at SED 195 and App. F.

88. The Knowledge Estimation Panel consisted primarily of teachers but also included teacher educators. Tr. at 2219. Members of the Knowledge Estimation Panel made judgments as to the performance of a minimally knowledgeable entry-level teacher for the purpose of calculating a proposed passing score. The means by which these judgments were made and compiled into a collective passing score is referred to in the educational measurement field as the Angoff method. Tr. at 2242–43; Ex. 801 at 34. This is the most commonly used method for setting a passing score and meets generally accepted professional standards. Tr. at 1487–88; SED Ex. 801 at 11–18. Dr. Faggen conducted extensive training of the panel members about how to make the knowledge estimation judgment, and panel members practiced and discussed such judgments prior to commencing this exercise. Tr. at 2243–53; SED Ex 20, G–6–15.

*17 89. The test question under consideration was projected onto a screen and panel members were asked to make an individual judgment about “the difficulty of individual test questions for persons who have the minimum level of knowledge and academic skills necessary for competent performance as a beginning teacher in New York State.” SED Ex 20, G–4; Tr. at 2242, 2244, 2253–55. Panel members were asked to make a judgment about the knowledge possessed by a minimally knowledgeable entry-level teacher because the Core Battery tests were designed for use as certification tests, not to assess knowledge or skills which might be learned on the job. Tr. at 2244, 2247–48; SED 11 at GUL 167. It was understood that some individuals taking the test would already have teaching experience when they sought certification in New York State. Tr. at 2247, 2256; SED 10 at GUL 137. ETS then aggregated the judgments of all panel members into a raw total score and converted that into a scaled score in order to reach a study value passing score for each test. Tr. at 2260–62; SED Exs. 4, 21.

90. Based on the analysis and conclusions of the validity study and the recommendation of SED staff, in December 1983 Commissioner Gordon Ambach selected the Core Battery tests to be New York State’s first teacher certification examinations. SED Ex. 11 at GUL 168; SED Ex. 14 at GUL. 141–42. As a result, persons who sought a provisional certificate as of September 2, 1984 in districts other than New York City and Buffalo were required to pass the Core Battery tests. SED Exs. 17, 19.

91. The cut (passing) scores recommended by the study for all three Core Battery tests were presented to the Commissioner, with a list of educational considerations that might affect his final selection of a passing score, including the chance that a candidate who actually possessed the knowledge and skills being assessed might nevertheless fail the test or that a candidate who did not, in fact, possess the knowledge and skills would pass (“the standard error of measurement”), and the supply and demand for new teachers. Tr. at 3050–53; SED Ex. 20, V.22–23; SED Ex. 4 at 3–4. The Commissioner asked the Teacher Education, Certification and Practice Board (TECAP) to review the proposed passing scores. SED Ex. 4. TECAP endorsed the use of the Core Battery tests for teacher certification in New York, observing “that the Core Battery tests reflect general knowledge applicable to all areas of certification,” and recommended that the passing scores be set at one standard error of measurement below the study value scores. SED Ex. 14 at SED 138–39; Ex. 4 at 3522–23. SED staff recommended that the passing scores be set at two standard errors of measurement below the study value scores. SED Ex. 16. In August 1984, the Commissioner set the cut score at

two standard errors of measurement below the study value score. Tr. at 2306–07, 3050–53; SED Ex. 4 at 3522–23; SED Ex. 16. New York continued to accept Core Battery test scores for certification for some teaching titles until September 1999. Tr. at 2318.

*18 92. In 1984 and 1985, ETS conducted an empirical job analysis to collect further validation evidence of the job relevance of the Core Battery tests to the job of public school teachers, regardless of subject matter or grade level. Tr. at 2107, 2197; SED Ex. 29 at E 410. Dr. Michael Rosenfeld testified about the methodology employed in this study. Tr. at 2139; SED Ex. 29. He has a Ph.D. in industrial psychology and measurement and has conducted over 40 job analyses relating to education and teacher certification. Tr. at 2095–96, 2100–01; SED Ex. 841. Dr. Rosenfeld is familiar with the APA/AERA/NCME standards and was at the time of the study. SEDExs. 841, 29 at E 419.

93. A job analysis is a systematic process for analyzing the knowledge and skills necessary to perform a job. Tr. at 2101, 2107; SED Exs. 29 at E 121–22. The job analyses provide the primary basis for defining the content domain. Although the job analysis techniques are comparable to those used in employment testing, the emphasis for certification is limited appropriately to knowledge and skills necessary to protect the public. Generally, knowledge and skills contained in a core curriculum designed to train people for the job or occupation are relevant. SED Ex. 121 at 64.

94. In conducting the Core Battery job analysis, Dr. Rosenfeld and his colleagues identified the core job functions of all teachers, regardless of grade level or subject matter taught; identified the content knowledge and skills deemed important by teachers and administrators for teachers to perform those core functions; and established the knowledge and skills necessary for all teachers upon entry into the profession. SED Ex. 29 at E 410; SED Ex. 800 at 36–37; Tr. at 3057–58. Three school districts were selected to participate in this intensive analysis of the job of a public school teacher: a large, urban Jersey City, New Jersey district; an unaffiliated group of rural and suburban schools in Costa County, California; and a suburban district in Albany, Georgia. Tr. at 2108, The districts were selected in part because they reflected a wide spectrum of geographical, demographic, socio-economic and educational characteristics. Tr. at 2108, 2145–46; SED Ex. 29 at E 411.

95. The researchers determined the knowledge and skills needed by all teachers through three means: a large-scale literature review of the teaching profession; interviewing

100 teachers in three diverse school districts; and conducting an extensive survey of teachers and administrators from all grade levels and teaching titles in those districts. Tr. at 2107–08; SED Ex. 29 at E 411.

96. They conducted an extensive search and review of published information about teachers’ duties, responsibilities and requirements, including job descriptions, job analyses and other related articles and books describing what teachers do and what they need to know to perform their jobs. SED Ex. 29 at E 411. Advisory Committees of teachers and teacher educators (including a principal from the Bronx) reviewed the information collected and drafted a survey describing the content knowledge and tasks necessary for all elementary and secondary school teachers. *Id.*; Tr. at 2110–11.

*19 97. The survey included 83 tasks performed by teachers and 59 knowledge areas, 34 of which were currently assessed on the Core Battery tests and 25 of which were under consideration for inclusion on the Core Battery tests. SED Ex. 29 at E 412–13; Tr. at 2113–23. One thousand eight hundred thirty-nine (1,839) teachers and 89 administrators in the three districts completed the surveys. SED Ex. 29 at E 413. They were asked, “How important is [this knowledge or skill] to YOUR professional functioning as a teacher?” SED Ex. 29 at C–1 (emphasis in original). Respondents were asked whether each knowledge or skill was required at entry level or could be learned on the job. Tr. at 2119; SED Ex. 29 at C–1. The goal was to assess the content knowledge and skills necessary upon entry into the profession; knowledge and skills which could be learned on the job did not need to be covered. Tr. at 2119–20.

98. Dr. Rosenfeld and his colleagues identified six core functions common to the jobs of all teachers. The importance of these functions was determined through a number of statistical techniques and confirmed for all subgroups (elementary and secondary teachers, sex, age, ethnicity, region of the country, years of teaching experience, grade level taught, and current teaching assignment). SED Ex. 29 at E 414–15 and App. J, Table 4; Tr. at 2131, 2142.

99. Through statistical means and professional judgments, the researchers then linked the core functions of the job of a teacher with the knowledge and skills assessed by the Core Battery which had been deemed necessary to perform the core functions effectively. Tr. at 2131–33; SED Ex. 29, VI–9, Table 19 and VI–15, Table 23.

100. The 1,839 teacher respondents agreed that all 34 content areas on the Core Battery were important, used by them personally in their jobs as teachers and necessary for effective performance as teachers. Tr. at 2131; SED Ex.

29, VI-5 and VI-6, App. K, Table 18. The empirical job analysis provided further validation evidence that the Core Battery tests assessed content knowledge and skills deemed necessary for public school teachers, regardless of subject matter taught or grade level taught. SED Ex. 29, VII; Tr. at 2139-41.

101. ETS issues Guidelines for the Proper Use of ETS Tests, which provide that ETS tests may not be used to adversely affect the credentialing of a fully licensed teacher. Tr. at 1963-64, 2263. SED Ex. 182 at SED 5305. Requiring temporarily or conditionally certified New York City teachers to take the Core Battery tests in order to become fully certified was consistent with ETS's guidelines for proper test use. Tr. at 1964-65, 1997, 2263.

102. Dr. William Mehrens testified that the validity evidence established that certification decisions made based on candidates' scores on the Core Battery tests were valid. Tr. at 3064-65; SED Ex. 800 at 32-37. Validity evidence established that test content was related to the job of a New York teacher. Tr. at 3058-59, 3043-50; SED Ex. 800 at 34-37. This evidence included the findings of the 1977 task force of New York educators that all teachers should be broadly educated in the liberal arts and sciences, the literature reviews, the work of the NTE Policy council and advisory committees, and the national survey of 3,000 educators, which informed the initial development of the Core Battery tests, the validation and standard setting study conducted exclusively for New York, ETS's 1986 empirical job analysis and the process followed by the Commissioner of Education in setting the passing score. Tr. at 3020-64; SED Exs. 800 at 33-37.

*20 103. After each test administration, ETS prepared statistical analyses of the Core Battery tests, including reliability data. SED Exs. 191-93, 196-97; Tr. at 1958-62, 1978-79. Dr. Mehrens testified that the reliability statistics indicate that New York's certification judgments based on the Core Battery tests were highly reliable. Tr. at 3053-56.

The LAST

104. In March 1988 Commissioner Sobol's Task Force on Teaching issued its "Blueprint for Learning and Teaching" recommending a statewide test of general liberal arts knowledge for provisional certification for all teachers, including those in New York City, and cautioning against reliance on temporary or emergency credentials, stating, "Anyone in the State using the title 'teacher' should possess a valid license." SED Ex. 35 at SED 1495-97.

105. The Blueprint for Learning and Teaching's recommendations for a test of liberal arts and sciences, as well as the similar recommendations in the 1977 task force report, were the basis for Regents' regulations establishing a test of the liberal arts and sciences for teacher certification. Tr. at 333-37. In May 1989 the Regents determined that individuals seeking provisional certification after September 1, 1993, must complete "a baccalaureate degree with a concentration in one of the liberal arts and sciences and college-level work in English, mathematics, science, social studies, and a language other than English," 8 NYCRR §§ 80.15(a)(1)(i) and 80.16(a)(1)(i), and pass a test of the liberal arts and sciences, which became the LAST, and a test of pedagogical knowledge, which became the Assessment of Teaching Skills-Written (ATS-W). Tr. at 1745; SED Exs. 35, 37, 42, 43. The Regents established additional requirements for permanent certification, to be met within five years, including: two years' teaching experience; a master's degree; passing content specialty tests specific to the individual's teaching title; and an Assessment of Teaching Skills-Performance (ATS-P). 8 NYCRR §§ 80.2(1), 80.2(m), 80.15(a) and 80.16(a). Tr. at 1710-24, 2315. The LAST is a single component of these requirements and tests, collectively called the New York Teacher Certification Examinations ("NYSTCEs"). Tr. at 1752.

106. The Commissioner issued requests for proposals ("RFPs") for all of the New York State Teacher Certification Examinations, requiring that they be closely aligned with the Regents' standards for students and the Regents' academic preparation requirements for teachers. Tr. at 273, 335-36; SED Exs. 45, 47, 51. ETS and National Evaluation Systems ("NES") submitted proposals for the test of liberal arts and sciences. Tr. at 1749-50. SED staff and a team of independent educators reviewed the two proposals. TECAP, which reviewed the proposals, determined that the knowledge and skills to be assessed on the test should have been acquired by the completion of college. Tr. at 3427, 3432.

107. In 1990, SED and NES entered into a contract for the development of the LAST. PL Ex. 74. Pursuant to that contract, SED did not pay NES any money to develop the LAST. Tr. at 274; PL Ex. 74. Instead, NES was to be paid through the administration of the test. Tr. at 274. SED fashioned this payment scheme because the State legislature would not budget the requisite funds to pay for test development. *Id.* NES has designed custom developed educational tests, including teacher certification tests, since 1972. NES-developed tests are used to certify teachers in New York, California, Texas, Illinois, Massachusetts, Arizona and New Mexico, among others. Tr. 699, 2589, 2840; Pl.s Ex. 253 at 84 n. 1.

*21 108. The team which worked on the development of the LAST included Dr. Gerald Freebome, Dr. Charles Mackey and Edith Hunsberger of SED, and NES's Dr. Paula Nassif, Dr. Scott Elliot, Dr. Homer Meade, Steve Weiss and Jeanne Clayton. Scott Elliot, who has an Ed.D. in measurement and assessment in research, was NES's Vice President of Client Services and project manager for the development of the NYSTCEs, including the LAST. Tr. at 384–86, 407, 423. Dr. Elliot is familiar with the joint AERA/APA/NCME standards. Tr. at 1181–82; SED Exs. 121,122.

109. NES set up two committees of New York educators whose responsibilities were primarily to review all proposed LAST content: the Content Advisory Committee (“CAC”) and the Bias Review Committee (“BRC”). Pl.Ex. 112 at N 2031–32; Tr. at 283–84, 780–88, 1129.

110. The CAC consisted of 22 teachers and education college faculty members with specialized expertise in liberal arts and sciences subjects. Tr. at 1130; SED Exs. 485, at N 2795; 357. Its charge was to review proposed LAST materials to ensure that the content was both free from bias and accurate and relevant to the job of a New York public school teacher. Tr. at 1129–30, 1104.

111. The BRC consisted of 22 New York public school teachers and education college faculty members; ten were Black, six Hispanic, three Asian, one Native American and two had disabilities. Tr. at 1130–31, 2599; SED Exs. 849; 485 at N 2795; 358. Its charge was to review proposed materials for all of the NYSTCEs, including the LAST, to ensure that the tests were free from bias. Pl.Ex. 112 at N 231; Tr. at 284, 1129–31, 1161, 2598–99, 2608. The BRC often broke up into smaller groups to review materials for individual NYSTCEs tests, including the LAST. Tr. at 1157, 786–89.

112. The approved draft framework was then circulated to New York public school teachers and New York education college faculty members for their review and comment in a job analysis survey. SED Exs. 401, 406, 408–10; Tr. at 1138–44. Although the court is suspicious of the actual numbers of surveys reportedly returned to NES, which suspicion is based on the fact that these surveys were destroyed by NES in contravention of the usual practices of test developers and on the complete inability of NES personnel called to testify at trial to remember specifics relating to who was involved in administering the survey (see discussion regarding lack of documentation *infra*, ¶¶ 143 to 146), the court does find that some sort of job analysis survey was done.

113. The average importance rating (on the 1 to 5 scale) for each of the 21 LAST objectives was 3.83 (a 4 being of

“great importance”). Tr. at 1149, 3076; SED Ex. 410 at N 536. The ratings for minority teacher respondents were comparable to that of white respondents. SED Ex. 410 at N 536; Tr. at 3076–77. The objective which received the highest rating from both teachers and faculty members was objective # 21: the ability to prepare, organize and write an essay. SED Ex. 410 at N 536; Tr. at 1152–53, 3076–78.

*22 114. As a result of the Regents' adoption of new learning standards in 1996, SED contracted with NES to review the NYSTCEs, including the LAST, to ascertain whether they closely tracked the Regents' academic expectations of students. Tr. at 1839–40; SED Exs. 512, 515, 542 at N 3762–63.

115. After the LAST's initial development, a number of national teaching associations issued standards for the credentialing of teachers, including the Interstate New Teacher Assessment and Support Consortium, the National Board for Professional Teaching Standards and the National Council for the Accreditation of Teacher Education. Tr. at 1838–39.

116. The LAST consists of 80 multiple choice questions, 64 of which are scored, and one essay question which assess a candidate's reading comprehension, writing skill, and analytical ability in the areas of mathematics, science and liberal arts. SED Ex. 485; SED Exs. 606–629.³ Initially there were approximately 350 items in the multiple choice data bank. Tr. at 1165–66, 2678. The essay portion of the test counted for 20 percent of a candidate's final score on the LAST. Tr. at 2766.

117. Test takers are sent a score report indicating performance by subarea level (e.g. math and science, historical and social scientific awareness, the essay question, etc.). Tr. 1881, 2798–99, 2801. Unlike with the Core Battery tests, where test takers could take the test at various locations around the country and elect whether or not to have their score forwarded to SED, because the LAST is a New York-specific test, NES administers the test and retains the scores of all LAST test takers. Upon registration, candidates are asked to provide information regarding their race or ethnicity and college at which they received their teacher preparation. Tr. at 2824.

118. As a group, minority candidates perform least well on the essay and the math and science sections of the test and best on the artistic expression and humanities section of the test. SED Exs. 496–502; Tr. at 2781–83, 2962. Of the three named plaintiffs in this suit, two were unable to pass the LAST because of the written component. SED Exs. 332–36, 342–43, 347. The last named plaintiff performed well on the essay section and passed the LAST accordingly. SED Ex. 349.

119. After each test administration, NES provides statistical analyses to SED about performance on that administration. Tr. at 2777–78.

120. At the end of each year, NES provides SED with annual statewide results. SED Exs. 496–503; Tr. at 2777–78. NES also provides teacher preparation colleges with statistics regarding the performance of their candidates for use in their reporting to the state and federal government. Tr. at 2793–97, 2800–01. Under both New York and federal law, education programs which have fewer than 80 percent of their graduates passing the teacher certification test risk losing their accreditation or federal funding. Tr. at 2821–22, 2958–65; 8 N.Y.C.R.R. § 52.21(b)(2)(iv)(b); Title II of the Higher Education Act Amendments, §§ 207 and 208.

*23 121. As shown below, neither NES nor SED properly identified the specific knowledge, skills and abilities that are critical to job performance and are likely to be present among the pool of people who meet the minimum qualifications of the job of New York public school teacher. Instead, NES presented test development committee members with test frameworks and a set of objectives that were never connected to the particular knowledge, skills, and abilities needed for successful performance of the job of New York public school teacher.

122. The Uniform Guidelines require:

[A] thorough job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them.... The work behaviors selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

29 C.F.R. § 1607.14(C)(2).

123. SED contemplated a “[s]equence of events” in which the first major work would be a “task analysis” of the tasks performed by a teacher. From the results of the task analysis, evidence was supposed to be gathered from which the objectives for the test would be drafted. Tr. at 114–15; Pl.Ex. 64 at 7.

124. No one knows whether this actually happened. Sanford Lake, an Assistant in Teacher Certification in SED’s Office of Teaching Initiatives, was not sure whether the 5 or 6 content domains of the test were arrived at after a task analysis was performed. Tr. at 161, 168–169. Edith Hunsberger, the Associate in Professional Examinations Development in SED’s Office of Teaching Initiatives during the relevant time periods and who was responsible for development of the LAST, acknowledged that the RFP envisioned the drafting of objectives only after a task analysis but testified that “[i]t may or may not have actually followed that order as it turned out.” Tr. at 701–702.

125. According to NES, the objectives and frameworks were to be based on materials to be collected from the field, on teacher curricula from colleges and New York public schools, and from interviews to be conducted with Deans from selected colleges. Pl.Ex. 107. There is no documentation describing any of the college materials collected or any information concerning the interviews. Tr. at 400–401; 405–408; 1187–1188.

126. In addition, the retained documents indicate that both the frameworks and the objectives were drafted before the collection of any curriculum materials or the holding of any interviews. As of March 12, 1991, NES had a draft of the frameworks and objectives, which it presented at a SED meeting. Pl.Ex. 105. The Plan for the Collection of College and University Curriculum Materials was initially approved after that meeting, on March 18, 1991. Pl.Ex. 107. NES has been unable to produce a draft of the objectives prior to April 18, 1991, and there is no way of knowing how the objectives were changed as a result of the March 12th meeting. Tr. at 1852–1853.

*24 127. According to NES officials, New York State public school teacher interviews were to be conducted, and a multi-page “teacher interview protocol” was prepared to be completed by the interviewers and used as the basis for drafting the frameworks. The only information in the record concerning these interviews shows that a total of eight people were interviewed. Pl.Ex. 129. There is no information as to how the interviewees were selected, who they were, or what they said. Pl.Ex. 289 at 125–126. No one knows what happened to the interview forms that were to be completed or whether they still exist. *Id.* at 126.

128. When people attending the March 12, 1991 meeting reviewed the objectives, they were given comment forms to fill out. Pl.Ex. 289 at 111. No one knows what happened to those completed forms. Pl.Ex. 289 at 112; Tr. at 1856. The Bias Review Committee, however, did recommend that it be reconvened in the fall of 1991 to

further review particular objectives. There is no evidence that this follow-up meeting was ever held. Pl. Exs. 120, 122; Tr. at 1866.

129. A 1991 LAST planning document stated that up to twenty-five people and six alternates would be selected for each of the bias and content review committees. There was a general understanding among SED officials that there should be between ten and twenty-five people on these committees who would review the test objectives and test items. Tr at 283; Pl. Exs. 279 at 25–26; 290 at 105–106.

130. SED solicited nominations for members of these committees and collected a batch of documents that was “three feet high,” and although SED was supposed to keep them, these documents were destroyed. Tr. at 792–93. Ms. Hunsberger acknowledged that there would be no way now to review who was considered, accepted or rejected. *Id.*

131. The Bias Review Committee was originally supposed to contain members “experienced in analyzing for test bias.” Pl.Ex. 97. That requirement was subsequently dropped; the application/nomination form for committee members contains no reference to experience in analyzing for test-bias. Pl.Ex. 97. Although notes were made and kept of discussions during committee meetings, only “summary notes” on the “consensus” of the committee members were retained. Tr. at 716–717, 720–7211.

132. With respect to the initial twenty-one test objectives and frameworks, only minor changes were made by the committees. Pl.Ex. 114. The frameworks ultimately approved for use were substantially and substantively identical to the initial draft presented by NES in March, 1991. Compare Pl.Ex. 126 with Pl.Ex. 114.

133. Although NES officials stated that content experts were consulted prior to the drafting of the objectives and that revisions in the objectives were made as a result of such input, no person has been identified nor has any of their work been described other than in extremely general terms. Tr. at 1853–1855, 1858. Scott Elliot, a project director and later the Vice President of Client Services for NES during the relevant time periods, could recall neither the name of any content expert nor any records that would identify them or what they did. Tr. at 394–396. The most information that Stephen Weiss, an editorial director for NES, could remember were the names of two individuals who “commented on the content that we had to find for the various objectives.” Tr. at 1858–59. Mr. Weiss testified that changes in the draft objectives were made by “national consultants,” but he did not identify who they were or how anyone could determine what changes were

actually made. Tr. at 1861–62.

*25 134. In its response to the request for proposal (“RFP”) issued by SED, NES stated that it had a standing “Equity Advisory Board” (“EAB”) which “generally reviews materials for inclusiveness and potential bias before they are presented to content or bias members” which “has been proven invaluable at detecting potential problems for Blacks, Latinos, women, Native Americans, Asian–Americans, and individuals with handicapping conditions.” Pl.Ex. 72 at 3556.

135. Regarding the EAB, no one from NES or SED who worked on the development of the LAST could identify a single person from the EAB with whom he or she met, or who did any work on the LAST, with one exception. Pl.Ex. 287 at 41. Dr. Paula Nassif was identified as the NES employee who worked most closely with the EAB. *Id.* She was able to name only one current member of the EAB and no members from previous boards. Pl.Ex. 288 at 22–23. She was unable to recall how many members there were on the board. PL Ex. 288 at 22. She identified Jeanne Clayton as the NES employee most likely to have “information” about the EAB. Pl. 288 at 23. Jeanne Clayton, a Senior Area Director for NES during the relevant times, testified that the EAB was asked to review materials on the LAST, but was unable to identify anyone on the EAB involved in this work. Pl. Ex, 289 at 66–69.

136. No one from NES or SED could articulate any specific action taken by the EAB beyond a vague suggestion to avoid portray Native Americans as victims in the test items. Tr. at 1859–61, 1068. No other witness could recall any other suggestion made by the EAB. Pl.Ex. 289 at 73.

137. The APA Standards require full documentation of the “procedures used to develop, review and try out items, and to select items from the item pool.” SED Ex. 122, Standard 3.7. In addition, item try-outs “should be documented” and the samples “should be as representative as possible of the population(s) for which the test is intended.” *Id.*, Standard 3.8.

138. One of SED’s witnesses, Dr. Carolyn Massad, who worked for ETS developing the Core Battery tests, confirmed that test developers follow the APA Standards with regard to document retention when creating test items. She stated that for each item created, ETS would retain a “flimsy”, which is “a piece of paper on which the item is typed. And it may be several sheets of paper attached together with all of the reviews and comments by everyone who has touched that test question and had something to say or do about it.” Tr. at 1982. ETS preserved these flimsies. *Id.*

139. NES represented that teachers, content specialists, and editors would be involved in the development of the items. PL Ex. 72 at 3551, 3778. However, neither Ms. Clayton nor any other witness could identify any person involved in drafting the items. PL Ex. 289 at 138. Mr. Weiss, who was in charge of item writing, testified that there is no way of determining who wrote items and what they were assigned to write. Tr. at 1869–70. No documentation shows the background, expertise or experience of any item writer.

*26 140. NES also represented in its response to the RFP that it would use a form for each draft test item containing information which required the item writer to record the author, date, objective, topic, item number and source, and included a large space for the drafting of each item. PL Ex. 72 at 3782. However, no such forms were produced. Mr. Weiss did not recall whether such a form was actually used in connection with the creation of the LAST. PL Ex. 290 at 92. Consequently, there is no way of determining what changes were made to any of the draft items. Tr. at 1870–71. The items for the LAST were not necessarily taken from specific source material but could have been drafted by an item writer based upon his or her own knowledge if he or she had expertise in that field. PL Ex. 290 at 120–21. There is no way of knowing what source material was used to create any test item. *Id.* at 125–126. Mr. Weiss testified that there were “item tryouts,” and that changes were made to the items as a result, but there is no documentation of this activity and no way of determining what was done in this regard. Tr. at 1871

141. When the new student learning standards were introduced in New York in the late 1990s, NES and SED purportedly conducted an investigation of the standards as they related to specific test items. However, there is no way of determining what was done, whether the items were changed or how any new items reflected the new learning standards. Tr. at 1883–1884.

142. NES represented in its response to the RFP that “[t]o insure a continuing stream of observations from the minority community, an *annual meeting* of the Minority Review Committee will be organized.” PL Ex. 72 at 3557 (emphasis in original). There is no evidence that this occurred. The disparity in passing rates on the LAST has increased over time. Tr. at 2693, 2854–59. No differential item analysis was undertaken. Tr. at 2674. The failure to conduct a differential item analysis violates APA Standard 7.3: “[W]hen credible research reports that differential item functioning exists ... test developers should conduct appropriate studies when feasible.” SED Ex. 122.

Pervasive Lack of Documentation

143. The original SED/NES contract contained a provision that required the retention of documents for a period of six years. The renewal contract, executed in 1995, contained the same provision. PL Ex. 74, App. A ¶ 10. Under these provisions all test development documents should have been retained and been available for review.

144. Despite the representations in NES’s documents, neither it nor defendants could identify: any person involved in collecting curricula, or any documents that would identify them, Tr. at 396:20–397; any educator interviewed by NES, Tr. at 400–401; any contacts with teachers in the field and field observations, Tr. at 405–408; any person involved in the creation and drafting of the frameworks, Tr. at 399, 437–38, 408–409; any documents which reflect the existence of the collected material or whether it is possible to determine what materials were used and how they were used, Tr. at 1843–48; whether materials were collected or interviews completed before the drafting of the objectives presented at the March 12, 1991 meeting, Tr. at 1851–53; the criteria that were used to develop the test’s frameworks and content, PL Ex. 289 at 113–16; or any person who was involved in drafting those frameworks, PL Ex. 289 at 101, 104–05.

*27 145. NES and SED retained insufficient documentary evidence to enable a person to reconstruct the process of developing the objectives. According to Mr. Weiss, this is the “wrong idea of the purpose of the process Our purpose was not to perform a scientific experiment in a way that made sure it could be replicated .” Tr. at 1855. Dr. Mehrens wishes that NES kept certain documents that they did not keep. Tr. at 3175. Specifically, Dr. Mehrens would have liked for NES to have kept sufficient documentation and data to allow someone to recreate the company’s validation results. Tr. at 3175–3176. He does not know why NES did not keep the test development documents. Tr. at 3193. Dr. Massad, a test developer at ETS, testified that “it is my responsibility as a professional” to document the test development process so that it could be reconstructed by a third party at some time in the future. Tr. at 1971.

146. This law suit is not the first time NES’s failure to retain documents has been criticized. The National Research Council (NRC), the research wing of the National Academy of Sciences (a federal agency comprised of the nation’s premier scientists, Tr. at 1339–40) compiled a report on teacher certification exams like the ones at issue in this case; this report, entitled “Testing Teacher Candidates,” was received into evidence. Tr. at

1322; Pl.Ex. 253. In the report, the NRC stated that it could not comment on the tests developed by NES, for the following reason:

Efforts to obtain sufficient technical information for the committee to evaluate the tests similar to what the committee received from ETS were unsuccessful for NES tests. As a result, ... the committee can make no statements about their soundness or technical quality.

The committee's inability to comment on NES-developed tests is significant. First, NES-developed tests are administered to very large numbers of teacher candidates [NY, MA, A2, CA, IL, TX, CO and OR all use NES-developed tests]. Second, the disclosure guidelines in the joint *Standards for Education and Psychological Testing* specify that "test documents (e.g. test materials, technical manuals, users guides, and supplemental materials) should be made available to prospective test users and other qualified persons at the time a test is published or released for use."

PL Ex. 253 at 84; Tr. at 1323. When the NRC requested the relevant materials from NES, NES responded that those materials were under the control of the client states and that requests must be made of those states. PL Ex. 253 at 84–85. When the NRC requested the relevant materials from the client states, they were told that their contracts with NES prohibited providing those materials to outsiders such as the NRC. Id at 85.

The LAST Cut Score

147. According to plaintiffs' expert, Dr. Landy, licensing and certification tests should be designed to prevent a specific harm. Tr. at 1310, 1360–1361. Dr. Mehrens agrees. Tr. at 3146.

*28 148. Dr. Mehrens is, however, unaware of any definition of harm specified or used by the State or NES in their development of the LAST. Tr. at 3148. Dr. Mehrens does not know whether the individuals who developed the LAST had an understanding of the concept of harm. Tr. at 3149. Committee members were never told that "harm prevention" was the objective of the LAST, nor were they informed of the type of harm or harms that the test was intended to reach. Tr. at 3149–3150. Committee members who rated the objectives for the LAST were never asked to look at or determine whether a test-taker's lack of knowledge about particular objectives or test items would cause harm. Tr. at 3149–3150. No committee members testified that the concept of harm was part of their analysis in their review of the frameworks, objectives, or items.

149. The passing score for the test recommended by the Content Advisory Committee after using the Modified Angoff method, was 48. Pl. Exs. 156; 163; 165; Tr. at 2839–2841. The Bias Review Committee performed the same exercise and recommended a passing score of 38. These recommendations were then taken to the State Commissioner who approved a passing score of 38. Tr. at 2841–2842. No one from SED or NES or any of the Committees ever performed an Angoff analysis on any of the other 270 items in the 350 item bank. Tr. at 2865–2867.

150. The Content Advisory Committee recommended increasing the passing score to 44. This was presented to the Commissioner, Tr. at 2659–66, 3308–15, who decided to adopt Jeanine Grinage's recommendation to increase the passing score incrementally from 38 multiple choice questions correct to 44 over three years, from 1998 to 2000. Tr. at 3308–15, 607–08, 612–13, 2666 –67; SED Exs. 588–89, 591–93, 782, 810. The Commissioner subsequently decided to make the final increase to 43 rather than 44. *Id.* At the time the passing score committee was determining whether the cutoff score should be raised to 44, it was given performance statistics showing that, if the cutoff score were to be set at 44, only 29% of African American test-takers and only 26% of Latino test-takers would pass, as opposed to 78% of White test-takers. Pl. Exs. 212, 214, 215, 217, 218.

151. Neither Ms. Grinage nor Commissioner Mills knew then, and they do not know today, of any study correlating teachers' performance on the LAST with the performance of students in the classroom. Tr. at 600, 3338. Although Ms. Grinage thought that such a study would be a "good idea," Tr. at 600, SED never performed such a study because "we had no money." *Id.*; Pl.Ex. 263. Commissioner Mills asked that a study be prepared, but that has not yet happened. Tr. at 3338–39.

152. According to Ms. Torres, who sat on one of the committees that recommended raising the cut score on the LAST, Tr. at 2913, the purpose of the LAST cut score is not to distinguish between minimally competent and incompetent teachers. Tr. at 2916. Dr. Mehrens and Dr. Landy agree that in fact this is precisely the purpose of a cutoff score on a licensing test. Tr. at 3215–3216; 1361.

*29 153. The court finds that there is insufficient documentation upon which to make a determination regarding the validity of the LAST for the uses to which it was put by defendants. Tr. at 1490.

PART III: PLAINTIFFS' PROPOSED ALTERNATIVES TO THE LAST

154. As purported alternatives to the LAST, plaintiffs advanced: (1) the use of “portfolio” assessments of teachers such as used in Connecticut, Tr. at 3509–42; and (2) the use of a “compensatory” model, in which a candidate who is unable to pass the LAST would be allowed to compensate for that inability with some other quality, such as a higher score on one of the other licensing tests or years of experience. Tr. at 3532–35.

155. Dr. Landy asserted that New York could employ “portfolio assessments” such as the system he claims is used by Connecticut, to replace the LAST. Tr. at 3514–15, 3532, 3542–43. He asserted that this “Connecticut scheme” for making certification decisions is more effective than the New York system Tr. at 3521; that Connecticut’s system would be related to the job of a New York teacher, Tr. at 3522; and that the use of the Connecticut scheme would have less of an adverse impact than the LAST. Tr. at 3526–32, 3542.

156. In fact, Connecticut does not use a portfolio review scheme instead of teacher testing; rather, Connecticut uses portfolios to assess the progress of its beginning teachers who have already been certified by passing both the Praxis I and Praxis II tests, and who are in their second year of teaching. Pl.Ex. 253 at 51, 153; Pl.Ex. 304 and 305 at 3. Dr. Landy admitted that Connecticut teachers have already passed two tests of academic knowledge and skills before undertaking a portfolio assessment in the area of their teaching specialty and that Connecticut does not use a portfolio model *instead of* tests, but uses it *in addition to* tests. Tr. at 3543–45. Dr. Landy testified that Praxis I may be waived as a requirement to enter education school if an applicant achieves a sufficient SAT or ACT score but that passing Praxis II is always a precondition to entry into teaching. Tr. at 3544.

157. Dr. Landy admitted that the statistical analyses that he relied upon to claim that the Connecticut portfolio review system has less adverse impact than tests was based on a population which has already passed these tests. Tr. at 3544. These analyses are not probative of a non-test-based alternative. The statistics set forth in the NRC report regarding the performance of minority candidates on the Praxis I and II tests suggests that the impact would be significantly greater. Pl.Ex. 253 at 108–09.

158. Dr. Landy suggested that a “compensatory model” would be an alternative to provide additional time to take the tests, under terms similar to those set forth in Commissioner Mills letter to Chancellor Crew. Tr. at 3535; Pl.Ex. 19. Dr. Landy did not explain how giving people more time to pass the LAST would be an “alternative” to having to take it eventually. This

alternative does not articulate a feasible, equally effective and less discriminatory way of assessing the same knowledge and skills as are measured by the LAST.

CONCLUSIONS OF LAW

*30 159. As explained *supra*, under Title VII disparate impact law, if plaintiffs establish a prima facie case of disparate impact, the burden shifts to defendants to show that the challenged practice is job related “and consistent with business necessity”. *Griggs v. Duke Power Co.*, 401 U.S. 424, 431–32 (1971); *Watson v. Forth Worth Bank & Trust*, 487 U.S. 977, 986 (1988); *Robinson v. Metro–North Commuter Railroad Co.*, 267 F.3d 147, 161 (2d Cir.2001). Where employment tests are challenged, the question is whether the test bears “a manifest relationship to the employment in question.” *Griggs*, 401 U.S. at 432 (citation and internal quotes omitted); *Watson*, 487 U.S. at 997–98.

160. To establish a prima facie showing in a disparate impact case, plaintiffs must

establish, by a preponderance of the evidence that the employer “uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.” 42 U.S.C. § 2000e–2(k)(1)(A)(i). To make this showing, a plaintiff must (1) identify a policy or practice, (2) demonstrate that a disparity exists, and (3) establish a causal connection between the two.

Robinson, 267 F.3d at 147. In this case, there is no dispute as to the first or third elements: the challenged policy is the use of the Core Battery and/or LAST, and plaintiffs were demoted because they were unable to pass them. The only question is whether plaintiffs demonstrated a disparity. The court finds that they did; accordingly, the court holds that plaintiffs have established their prima facie case.

161. The burden having shifted to defendants, the court must decide whether they have demonstrated that the Core Battery and the LAST are job related. Plaintiffs urge the court to hold defendants to the disparate impact standard enunciated by the Second Circuit in the 1970s and 1980s in relation to challenges to standardized testing namely, formal validity. See *Guardians Ass’n of New York City Police Dep’t, Inc. v. Civil Serv. Comm’n*, 630 F.2d 79, 95–105 (2d Cir.1980), *cert. denied*, 452 U.S. 940 (1981); *Guardians Ass’n of New York City Police Dep’t, Inc. v. Civil Serv. Comm’n*, 633 F.2d 232 (2d Cir.1980). Under this standard, while defendants would prevail with

respect to the Core Battery, plaintiffs would emerge triumphant with respect to the LAST.

162. Unhappily for plaintiffs, however, the Supreme Court lowered the bar for defendants in disparate impact suits. In *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977 (1988), the Court explained:

Our cases make it clear that employers are not required, even when defending standardized or objective tests, to introduce formal “validation studies” showing that particular criteria predict actual on-the-job performance. In [*New York City Transit Authority v. Beazer*, 440 U.S. 568 (1979)], for example, the Court considered it obvious that legitimate employment goals of safety and efficiency permitted the exclusion of methadone users from employment with the New York City Transit Authority; the Court indicated that the manifest relationship test was satisfied even with respect to non-safety-sensitive jobs because those legitimate goals were significantly served by the exclusionary rule at issue in that case even though the rule was not required by those goals.

*31 *Id.* at 998 (citations and internal quotation marks omitted). In the instant case, defendants have shown that the LAST is manifestly related to the legitimate educational goals enunciated by SED.

163. The court concludes that the LAST is job-related, despite defendants’ failure to demonstrate its formal validity, because of the coincidence of three factors: (1) the importance given to the ability to write an essay by those education professionals surveyed by NES (*see supra*, ¶ 113); (2) the weight of the essay writing portion of the test (*see supra*, ¶ 116); and (3) the fact that the majority of plaintiffs would have passed the LAST but for the essay writing portion *see supra*. ¶ 118). The education professionals surveyed by NES stated that of all the skills required of teachers, the ability to write a coherent essay was most important. For this reason, the essay portion of the test was given more weight than any other, a full

twenty percent. Plaintiffs, however, were brought low by this section; the evidence introduced at trial demonstrated that the majority of them performed worse on the essay portion than on any other section. The named plaintiffs who failed the LAST did so because of their poor performance on the essay portion; the named plaintiff who passed did well on that section.

164. It should go without saying that New York City teachers should be able to communicate effectively in both spoken and written English. Teachers who are unable to write a coherent essay without a host of spelling and grammar errors may pass on that deficiency to their students, both in commenting upon and grading the work they turn in. Defendants’ decision to exclude those who are not in command of written English is in keeping with the legitimate educational goal of teaching students to write and speak with fluency. The court concludes that defendants have met their burden and shown that the LAST bears a manifest relationship to teaching.

165. Plaintiffs have failed to offer a cost effective, practical alternative to the tests used by defendants in certifying teachers. See *Watson*, 487 U.S. at 998 (“Factors such as the cost or other burdens of proposed alternative selection devices are relevant in determining whether they would be equally as effective in serving the employer’s legitimate business goals”). Although this court was impressed by the portfolio method of certification used by the state of Connecticut, the evidence demonstrated that it was used in addition to, not instead of, testing. Accordingly, the evidence did not clearly demonstrate that instituting Connecticut’s certification process would have less of a disparate impact. See *Robinson*, 267 F.3d at 161. The evidence adduced regarding a compensatory model alternative was simply too thin. See *New York City Envtl. Justice Alliance v. Giuliani*, 214 F.3d 65, 72 (2d Cir.2000).

Conclusion

In light of the foregoing, the court finds for defendants. Accordingly, it need not address the Board’s crossclaims against SED.

*32 SO ORDERED.

Footnotes

¹ The definition of what inference is being made is further refined with reference to the issue of harm at ¶ 147, *supra*.

² SED Ex. 329 was sealed by order of this court to protect the secrecy of the test items.

³ SED Exs. 606–629 were sealed by order of this court to preserve the secrecy of the test items.