

TESTOCRACY: THE UNDEMOCRATIC SYSTEM OF STANDARDIZED TESTING IN THE UNITED STATES

*By James Naughton**

Education has been a contentious issue since Socrates drank hemlock for corrupting the youth of ancient Greece. This problem plagues us to this day, and without a solution to the issues around education and especially educational testing, we are doomed to die the same death as Socrates. To avoid these end of the proverbial educational days, we need to look at our country's history of education and radically redefine how we assess students' progress—we must abolish standardized tests. While scholars have written numerous papers on this topic, none have parsed the data and revealed the flaws inherent in a crucial test for eighth and ninth graders, the PSAT 8/9. This article will take a deep statistical dive into the PSAT 8/9 to prove that the College Board who administers the PSAT 8/9, SAT, and a suite of other assessments have led education astray. The importance of developing rigorous, equitable testing has never been so important in the days of reckoning for racial justice. This article will offer several alternatives to standardized tests, most of them novel.

INTRODUCTION

The concept of legal, compulsory education took root early in the history of the United States. Massachusetts was the first colony to adopt compulsory education laws with ramifications for student non-attendance.¹ It did not take long for other New England colonies to follow suit and pass their own compulsory attendance acts. Connecticut passed its own code in 1655² and the

* James Naughton received his J.D., *magna cum laude* from Loyola University Chicago School of Law in 2018. Mr. Naughton received his Bachelor of Arts degree in History, *summa cum laude* from Loyola University Chicago. He currently serves as an Assistant States' Attorney in Cook County, Illinois. The author dedicates this article in memory of Nick Mizenberger—one of the best skateboarders this author had the pleasure of knowing—may he skate in peace forever. Until next time, Nick.

¹ MAX FARRAND, *THE LAWS AND LIBERTIES OF MASSACHUSETTS* 42 (Harvard Univ. Press 2d ed. 1929) (discussing the Old Deluder Satan Act and the five pound per annum charged to a town by the Royal Charter if the town did not establish a school).

² M.S., *NEW HAVEN'S SETTLING IN NEW ENGLAND* (London, 1656), reprinted in CHARLES

Plymouth Colony in 1671.³ As education was primarily seen as a state's prerogative in 1900, half of the existing states had compulsory education laws.⁴ By 1930, every U.S. state had compulsory education laws in place.⁵ The most recent data released by the U.S. Department of Education's National Center for Education Statistics shows that average daily attendance ("ADA") in all fifty states for students in elementary and secondary schools now hovers at approximately ninety percent.⁶

As schools developed in the United States, so did the idea for a standardized system of testing.⁷ Before there was universal compulsory education, there was universal standardized testing. This testing began in 1901 after the creation of the College Entrance Examination Board.⁸ Between 1900 and 1932, more than 1,300 achievement tests were on the market during the nescient years of compulsory education laws, including ninety-two tests for secondary education students.⁹ In 2015, the average student in America's public schools took "some 112 mandatory standardized tests between pre-kindergarten and the end of 12th grade – an average of about eight a year . . . [t]hat eats up about 20 and 25 hours every school year."¹⁰ One researcher said of this system of standardized testing that "I witnessed a systemic violence there [at the school], an institutional brutality that was hurled upon students, families, teachers, and administrators."¹¹ One educator pointed out the stress of

JEREMY HOADLY, RECORDS OF THE COLONY OR JURISDICTION OF NEW HAVEN 583–88 (1858).

³ See COUNCIL FOR NEW ENGLAND, UNITED COLONIES OF NEW ENGLAND, NEW PLYMOUTH COLONY & WILLIAM BRIGHAM, THE COMPACT WITH THE CHARTER AND LAWS OF THE COLONY OF NEW PLYMOUTH: TOGETHER WITH THE CHARTER OF THE COUNCIL AT PLYMOUTH, AND AN APPENDIX, CONTAINING THE ARTICLES OF CONFEDERATION OF THE UNITED COLONIES OF NEW ENGLAND, AND OTHER VALUABLE DOCUMENTS 270–71 (1836).

⁴ See MICHAEL S. KATZ, A HISTORY OF COMPULSORY EDUCATION LAWS 17–18 (1976), <https://theconversation.com/federal-role-in-education-has-a-long-history-74807> [<https://perma.cc/Z7HM-YWHR>].

⁵ Dustin Hornbeck, *Federal Role in Education Has a Long History*, CONVERSATION (Apr. 26, 2017, 9:51 PM EDT), <https://theconversation.com/federal-role-in-education-has-a-long-history-74807> [<https://perma.cc/3WF7-VJLD>].

⁶ NAT'L CTR. FOR EDUC. STAT., AVERAGE DAILY ATTENDANCE (ADA) AS A PERCENTAGE OF TOTAL ENROLLMENT, SCHOOL DAY LENGTH, AND SCHOOL YEAR LENGTH IN PUBLIC SCHOOLS, BY SCHOOL LEVEL AND STATE: 2007-08 AND 2011-12 (2012), https://nces.ed.gov/programs/digest/d18/tables/dt18_203.90.asp [<https://perma.cc/WX4P-MQ9G>].

⁷ See generally NAT'L EDUC. ASS'N, *History of Standardized Testing in the United States* (June 25, 2020), <https://www.nea.org/professional-excellence/student-engagement/tools-tips/history-standardized-testing-united-states> [<https://perma.cc/G9QD-FPP4>].

⁸ *History of the SAT: A Timeline*, PUB. BROAD. SERV., <https://www.pbs.org/wgbh/pages/frontline/shows/sats/where/timeline.html#:~:text=In%201901%2C%20the%20first%20College,questions%20from%20the%201901%20test.&text=A%20French%20psychologist%2C%20Alfred%20Binet,that%20could%20measure%20one's%20intelligence> [<https://perma.cc/399J-F2KZ>].

⁹ NAT'L EDUC. ASS'N, *supra* note 7.

¹⁰ Valerie Strauss, *Confirmed: Standardized Testing Has Taken Over Our Schools. But Who's To Blame?*, WASH. POST (Oct. 24, 2015), <https://www.washingtonpost.com/news/answer-sheet/wp/2015/10/24/confirmed-standardized-testing-has-taken-over-our-schools-but-whos-to-blame/> [<https://perma.cc/W35V-E3Q6>].

¹¹ M. FRANCYNE HUCKABY, *An Invitation, in PUBLIC EDUCATION/PARTICIPATORY DEMOCRACY: AFTER NEOLIBERALISM* (citing TERRY JO SMITH, DEFENDING PUBLIC EDUCATION

parents raising children “in the age of what we could call ‘testocracy’ – the test-and-punish corporate education reform tsars whose interests are served by the proliferation of high-stakes standardized bubble tests.”¹²

The first Section of this Article reviews the federal and state history of standardized testing proliferation and conversations amongst educational stakeholders on the issue. Then, in Section two, the Article dives into a review of the PSAT 8/9 data presented by the College Board and argues that statistically speaking, the data put out is worthless. Third, this Article uses social science research to demonstrate that standardized tests do not account for extraneous variables. Once other factors are accounted for, standardized tests are correlative at best. Finally, this Article suggests the abolition of standardized testing and offers an alternative method of collecting student progress through formative and summative assessments.

I. THE PROLIFERATION OF STANDARDIZED TESTS: HOW DID WE GET HERE?

As the Massachusetts Bay Colony was the first to pass a compulsory attendance act, it may be no surprise that the Father of the Standardized Test, Horace Mann, hailed from Massachusetts.¹³ Mann served as Massachusetts’s Secretary of Education from 1837 to 1848, when oral examinations were the primary way school teachers assessed students.¹⁴ In 1843, Mann traveled across Europe to inspect their educational systems and came back convinced that written examinations would prove the superior tool for assessment.¹⁵ In 1845, Mann and members of his board of education created and administered written examinations to students in Boston.¹⁶ Similar to today, the State Board of Education used standardized test results to chastise teachers while teachers protested that the standardized test had little to do with what they taught school children in the classroom.¹⁷ From the inception of standardized testing, education administrators used the results to sort and classify students and monitor school effectiveness.¹⁸

A mere sixty-six years after Mann’s first standardized tests, Fredrick J. Kelly created a truly state-wide standardized assessment in the state of

FROM CORPORATE TAKEOVER 41 (Todd Alan Price, John Duffy & Tania Giordani eds., 2013), <https://scalar.usc.edu/works/publiceducation/testing?path=theory--themes> [https://perma.cc/34SU-NDGP].

¹² JESSE HAGOPIAN, MORE THAN A SCORE: THE NEW UPRISING AGAINST HIGH-STAKES TESTING 8 (Haymarket Books 2014).

¹³ LEHIGH UNIV. COLLEGE OF EDUC., *History of Standardized Testing* (Oct. 18, 2013, 1:30 PM), <https://ed.lehigh.edu/news-events/news/history-standardized-testing> [https://perma.cc/4YCM-XR2C].

¹⁴ *Id.*

¹⁵ *Id.*

¹⁶ *Id.*

¹⁷ *Id.*

¹⁸ OFF. OF TECH. ASSESSMENT, TESTING IN AMERICAN SCHOOLS: ASKING THE RIGHT QUESTIONS, OTA-SET-519 1, 107 (Feb. 1992).

Kansas.¹⁹ During these sixty-six years, Mann's idea spread like wildfire from the East to the West Coast, and by the time Kelly's multiple-choice test—The Kansas Silent Reading Test—came about, almost every state had some version of standardized, written testing.²⁰ During the midst of the standardized testing craze, the original Department of Education was created in 1867 to help establish effective school systems and collect data about the nation's schools.²¹

As education had been the states' prerogative since the founding of the colonies, there was natural resistance to a federal department of education as states and standardized test-makers were now the tycoons of education.²² After hardly a year in existence, the U.S. Department of Education was demoted in an appropriation bill in 1868, given three employees, and tucked away into the Department of the Interior.²³ The U.S. Department of Education may have been tucked away in 1868, but the U.S. Department of War—now the U.S. Department of Defense—was at the forefront of standardized testing.

In 1905, French psychologist Alfred Binet began to develop a standardized test to measure intelligence.²⁴ Binet's work would be expounded upon by Stanford Professor Lewis Terman, who took Binet's intelligence quota ("IQ") test and created a standardized test for large-scale IQ testing in 1912 known as the Stanford-Binet IQ test.²⁵ On July 28, 1914, World War I broke out, starting the Great War and the Great Test.²⁶ This test, called the Alpha and Beta test by the U.S. military, was used to screen approximately 1.75 million draftees during World War I.²⁷ As World War I wound to a close in 1918, Stanford's tests were just beginning to take root. In 1930, Stanford's newly published "Stanford Achievement Test," now known as the "SAT," had sold over two million copies per year to schools since 1918.²⁸ In the twelve years since the advent of his Stanford-Binet test, Stanford sold over twenty-four million copies of his SAT to state school systems.²⁹

The College Board, founded in 1901, adopted the SAT as the national test

¹⁹ *History of Standardized Tests*, PROCON.ORG (Dec. 7, 2021), <https://standardizedtests.procon.org/history-of-standardized-tests/> [https://perma.cc/D2KZ-MQTT].

²⁰ OFF. OF TECH. ASSESSMENT, *supra* note 18, at 254.

²¹ See generally An Act of Mar. 2, 1867, ch. 158, 14 Stat. 343 (establishing the Department of Education).

²² *An Overview of the U.S. Department of Education*, U.S. DEP'T OF EDUC. (Sept. 2010), <https://www2.ed.gov/about/overview/focus/what.html> [https://perma.cc/YL59-WSDM].

²³ Kevin R. Kosar, *Department of Education Abolition Act of 1868*, FED. EDUC. POL'Y HIST. (Sept. 10, 2015), <https://federaleducationpolicy.wordpress.com/2015/09/10/department-of-education-abolition-act-of-1868-2/> [https://perma.cc/8Z4S-RMCZ].

²⁴ NAT'L EDUC. ASS'N, *supra* note 7.

²⁵ *Id.*

²⁶ *Id.*

²⁷ Daphne Martschenko, *IQ Tests Have a Dark, Controversial History — But They're Finally Being Used for Good*, BUS. INSIDER (Oct. 11, 2017, 12:34 PM), <https://www.businessinsider.com/iq-tests-dark-history-finally-being-used-for-good-2017-10> [https://perma.cc/8RYH-ZE8L].

²⁸ NAT'L EDUC. ASS'N, *supra* note 7.

²⁹ JOSEPH F. KETT, *MERIT: THE HISTORY OF A FOUNDING IDEAL FROM THE AMERICAN REVOLUTION TO THE 21ST CENTURY* 129 (2013).

in 1926 for college admissions.³⁰ As enthusiasm for standardized testing began to falter, the pushback against the new test regime began to pick up steam.³¹ Standardized testing enthusiasm has held steady, and still in 2019, half of all U.S. states required that students take the SAT.³² The United States, again spurred by the potential for war, saw leaps in educational funding during the 1950's Cold War with the Soviet Union.³³ In 1979, Congress passed the Department of Education Organization Act, which elevated the U.S. Department of Education to its own Department with its own presidential cabinet position.³⁴

The states, once resistant and protective of their standardized tests, ultimately lost the war to the federal government. President Lyndon B. Johnson fired the first volley in 1965 with Congress's passage of the Elementary and Secondary Education Act ("ESEA").³⁵ President Johnson's ESEA was relatively mundane—providing resources to schools, funds for professional development, and resources to create educational programs.³⁶ Then, President George W. Bush, Jr., signed into law the now infamous No Child Left Behind Act ("NCLBA") which amended the 1965 ESEA and established state testing requirements to assess "adequate yearly progress" of students.³⁷ On December 10, 2015, the Every Student Succeeds Act ("ESSA") was signed into law by President Barack Obama, replacing the NCLBA.³⁸ President Obama touted the new law as a radical departure from former President Bush's NCLBA—a claim that was only partially true.³⁹ While ESSA removed the federal monetary penalties attached to student success in the NCLBA and states can now consider school-quality factors, the simple fact is that students are still required

³⁰ NAT'L EDUC. ASS'N, *supra* note 7.

³¹ *Id.*

³² Catherine Gewertz, *Which States Require Students to Take the SAT or ACT*, EDUC. WEEK (Feb. 15, 2017), <https://www.edweek.org/teaching-learning/which-states-require-students-to-take-the-sat-or-act> [<https://perma.cc/JQ82-5MSH>].

³³ See Amanda Pineo, *Cold War in the Classroom: Effects of the Onset of the Cold War on Public Education*, VIRTUAL COMMONS-BRIDGewater STATE UNIV. 1, 14 (May 8, 2018), http://vc.bridgew.edu/honors_proj/273 [<https://perma.cc/5U5H-MN7X>]; see also Dave Roos, *How the Cold War Space Race Led to U.S. Students Doing Tons of Homework*, HISTORY (Aug. 13, 2019), <https://www.history.com/news/homework-cold-war-sputnik> [<https://perma.cc/4APJ-C2B4>].

³⁴ See Department of Education Organization Act, Pub. L. No. 96-88, 93 Stat. 668 (1979) (codified as amended at 20 U.S.C. § 3411).

³⁵ See Elementary and Secondary Education Act (ESEA) of 1965, Pub. L. No. 89-10, 79 Stat. 27 (codified as amended at 20 U.S.C. ch. 70).

³⁶ *Id.*

³⁷ No Child Left Behind Act (NCLB) of 2001, Pub. L. No. 107-110, 115 Stat. 1425, 1446–47 (2002) (codified in scattered sections of 15 U.S.C., 20 U.S.C., 42 U.S.C. & 47 U.S.C.); see also LaTefy Schoen & Lance D. Fusarelli, *Innovation, NCLB, and the Fear Factor: The Challenge of Leading 21st-Century Schools in an Era of Accountability* 22 EDUC. POL'Y 181, 182 (2008) (discussing the No Child Left Behind Act more in-depth).

³⁸ U.S. DEP'T OF EDUC., *Every Student Succeeds Act (ESSA)* (Dec. 10, 2015), <https://www.ed.gov/essa?src=ft> [<https://perma.cc/D82S-4C72>].

³⁹ *Id.*

to take a federally mandated test that sets target goals.⁴⁰

Just as teachers and students pushed back in 1845 against Horace Mann's standardized testing in Boston, there exists a current movement that continues to push back against high-stakes, standardized testing.⁴¹ While teachers, students, and parents may all have their own personal reason for being abolitionists of standardized testing, there is a tendency to focus around the fundamental unfairness to teachers, the instructional time taken away from students, the inability to account for all the variables that make up a child, and the disadvantages to students of color and students with disabilities. The following Section dives into discussions amongst educators, students, and teachers about standardized tests.

A. *The Lost Voices of Educators and Students*

There is a history of lamentation over the use of standardized testing in the United States. Almost as quickly as standardized tests appeared on the world stage, so did criticism of their efficacy, overuse, and tendency to crush creativity in abolitionists' opinions.⁴² One of the strongest, earliest voices of the anti-standardized test movement was American philosopher John Dewey. John Dewey was one of the principal figures of the Progressive Educational Movement in the United States⁴³ who believed in two fundamental educational principles:

(1). *Respect for diversity*, meaning that each individual should be recognized for his or her own abilities, interests, ideas, needs, and cultural identity, and (2). the development of *critical, socially engaged intelligence*, which enables individuals to understand and participate effectively in the affairs of their community in a collaborative effort to achieve a common good. These elements of progressive education have been termed "child-centered" and "social reconstructionist" approaches.⁴⁴

Inevitably, the Progressive Education Movement's approach would

⁴⁰ Lindsay Jones, *The Difference Between the Every Student Succeeds Act and No Child Left Behind*, UNDERSTOOD (2015), <https://www.understood.org/en/school-learning/your-childs-rights/basics-about-childs-rights/the-difference-between-the-every-student-succeeds-act-and-no-child-left-behind> [https://perma.cc/3JAV-CZ5L].

⁴¹ Derek Cantù, *Lawmakers Push to Waive Standardized Testing in 2021, Despite Feds' Instance*, NAT'L PUB. RADIO (March 15, 2021), <https://www.nprillinois.org/statehouse/2021-03-15/lawmakers-push-to-waive-standardized-testing-in-2021-despite-feds-insistence> [https://perma.cc/39KT-7LJE].

⁴² See generally NAT'L EDUC. ASS'N, *supra* note 7.

⁴³ Alberto Piedra, *The Tragedy of American Education: The Role of John Dewey*, INST. OF WORLD POLS. (Feb. 1, 2018), <https://www.iwp.edu/articles/2018/02/01/the-tragedy-of-american-education-the-role-of-john-dewey/> [https://perma.cc/YMD3-RL9G].

⁴⁴ *A Brief Overview of Progressive Education*, UNIV. OF VT. DEP'T OF EDUC. (Jan. 30, 2002), <https://www.uvm.edu/~dewey/articles/proged.html> [https://perma.cc/48GY-HQKC].

come to loggerheads with the proponents of standardized testing such as Horace Mann. Dewey, writing of the test-taking industry at the turn of the twentieth century, stated, “our mechanical, industrialized civilization is concerned with averages, with percents. The mental habit which reflects this social scene subordinates education and social arrangements based on average gross inferiorities and superiorities.”⁴⁵ While Dewey built on Mann’s ideas and agreed with many of his radical educational policies, Dewey would never side with Mann on standardized testing.⁴⁶ Dewey was about a century ahead in his criticism attacking standardized tests based on the variables of economic inequalities and social injustices at the turn of the twentieth century.

Following close on the heels of the SAT came E.F. Lindquist’s creation of the Iowa Testing Program in 1929.⁴⁷ Throughout the 1930s, the Iowa Testing Program and the New York State Regent’s Test, used in 1876⁴⁸ for the first time, were met with fierce opposition from educators.⁴⁹ It is best to break down the opposition to standardized testing into three categories: 1) distortion; 2) waste; and 3) (mis)classification.⁵⁰ The first of these categories, distortion, boils down to the idea that standardized testing distorts or disrupts the regular curriculum of a classroom teacher.⁵¹ As one teacher wrote in a 1927 *New York Times* article, it is “quite possible to drill for an examination and to pass a large number of pupils with high ratings without giving any breadth of outlook or grasp of underlying principles.”⁵² Fifty years after this article was published, Albert Shanker, the President of the American Federation of Teachers, stated that “schools are devoting more and more time teaching kids strategies for filling in blanks and choosing answers to multiple-choice questions. This destroys much of the value of these tests, which only tell you something if they are an independent measure of what the student knows.”⁵³ This Article expounds in later sections on Shanker’s issue with the independent measurement of knowledge and uncontrolled variables but suffice it to say that

⁴⁵ NAT’L EDUC. ASS’N, *supra* note 7.

⁴⁶ See generally Audrey Cohan & Charles F. Howlett, *The Historic Link between Horace Mann and John Dewey in Support of Public Schooling: A Lesson in Democracy*, MOLLOY COLL. (2019), <https://teachingsocialstudies.org/2019/07/06/the-historic-link-between-horace-mann-and-john-dewey-in-support-of-public-schooling-a-lesson-in-democracy/> [<https://perma.cc/547Y-DYHL>].

⁴⁷ ETHAN HUTT & JACK SCHNEIDER, A HISTORY OF ACHIEVEMENT TESTING IN THE UNITED STATES OR: EXPLAINING THE PERSISTENCE OF INADEQUACY 14 (2018), https://www.holycross.edu/sites/default/files/educ/a_history_of_achievement_testing.pdf [<https://perma.cc/AC9U-PZKB>].

⁴⁸ Joseph Spector, *Regents Exam in New York: Why the Century-old test may be scrapped*, DEMOCRAT & CHRON. (July 25, 2019, 1:09 PM), <https://www.democratandchronicle.com/story/news/politics/albany/2019/07/25/regents-exams-new-york-how-century-old-test-may-scrapped-ended-elias-rosa/1825691001/> [<https://perma.cc/EKE2-B7Q8>].

⁴⁹ HUTT & SCHNEIDER, *supra* note 47, at 3.

⁵⁰ *Id.* at 2.

⁵¹ *Id.* at 6.

⁵² JACK SCHNEIDER, BEYOND TEST SCORES: A BETTER WAY TO MEASURE SCHOOL QUALITY 26 (Harvard Univ. Press 2017).

⁵³ HUTT & SCHNEIDER, *supra* note 47, at 6.

Shanker stood on solid ground.

The second criticism—waste—focuses on the lost instructional time as educators and students jettison their regular curriculum “substituting test preparation for instruction.”⁵⁴ In 1930, Dr. Henry Linville, then president of the New York Teachers Union, stated that the Regents tests were a “continuing waste of childhood that is appalling to contemplate.”⁵⁵ According to research conducted by the *Washington Post*, the time consumed by standardized testing alone—not including preparation—takes up 2.34% or 4.22 days of a typical 180 day school year.⁵⁶ Research has also shown that the average teacher spends fourteen days preparing students for state-mandated exams and twelve days for district-mandated exams.⁵⁷ Educators and students are pulled away from content learning for roughly thirty days out of an average 180 day school year to prepare for standardized tests.⁵⁸ This is 16.7% of the entire school year.⁵⁹

The final criticism is that of misclassification. The thrust of the misclassification is that standardized tests often misclassify students, schools, and during the regime of NCLBA—teachers.⁶⁰ In the same 1927 *New York Times* article, a teacher pointed out that test results are “used to make invidious comparisons.”⁶¹ The misclassification argument intensified during the NCLBA era with no less than five federal court lawsuits challenging a teacher’s classification based on their student’s test scores.⁶² Aside from the impact on teacher classification, students are often tracked or, simply put, placed into lower-level courses based on their standardized test scores.⁶³ These attempts to create homogenous groups of students almost always fail. As educational researcher Kevin G. Welner found, most tracked classes in writing, e.g., remedial, regular, enriched, and advanced, were far from homogenous in his

⁵⁴ *Id.*

⁵⁵ *Id.* at 6.

⁵⁶ See Lyndsey Layton, *Study Says Standardized Testing is Overwhelming Nation’s Public Schools*, WASH. POST (Oct. 24, 2015), https://www.washingtonpost.com/local/education/study-says-standardized-testing-is-overwhelming-nations-public-schools/2015/10/24/8a22092c-79ae-11e5-a958-d889faf561dc_story.html [<https://perma.cc/4JAJ-9226>]; see also Caroline Waldman, *TESTING OVERLOAD?: New Study Finds U.S. Students Are Spending Time Taking Redundant Tests*, 15 PUB. EDUC. POL’Y & PROGRESS 20 (Nov. 10, 2015).

⁵⁷ Erik Robelen, *Testing and Test Prep: How Much Is Too Much?*, EDUC. WRITERS ASS’N (June 3, 2016), <https://www.ewa.org/blog-educated-reporter/testing-and-test-prep-how-much-too-much#:~:text=On%20average%2C%20teachers%20estimate%20spending,good%2C%20some%20not%20so%20good> [<https://perma.cc/Y423-D2QU>].

⁵⁸ *Id.*

⁵⁹ *Id.*

⁶⁰ HUTT & SCHNEIDER, *supra* note 47, at 4.

⁶¹ *Id.* at 7.

⁶² See *Wash. Teachers’ Union v. D.C. Pub. Schs.*, 207 A.3d 1143 (D.C. 2019); see also *Lederman v. King*, 54 Misc. 3d 886 (N.Y. Sup. Ct. 2016); see also *Trout v. Knox Cnty. Bd. of Educ.*, 163 F. Supp. 3d 492 (E.D. Tenn. 2016); see also *State ex rel. Stapleton v. Skandera*, 346 P.3d 1191 (N.M. Ct. App. 2015); see also *Wash. Teachers’ Union v. D.C. Public Schools*, 77 A.3d 441 (D.C. 2013).

⁶³ HUTT & SCHNEIDER, *supra* note 47, at 7.

observational study.⁶⁴ Welner points to another flaw in standardized testing as a possible explanation—while students’ mean scores were different—the overlap in the range of scores was remarkable.⁶⁵ In other words, the test did not consider this vast range of scores because the class was designed for a somewhat narrow pocket of mean scores. To offer a hypothetical, say a standardized examination was 100 points—you will have classrooms made of clusters of students whose mean scores were the same while you may have students who scored closer to the bottom twenty percent all the way up to the top twenty percent.

Parents, too, have taken up the mantle against standardized tests. In her collection of essays from educators, students, and parents discussing standardized testing, educator Jesse Hagopian, interviewed Helen Gym, a teacher and a parent of a student in Philadelphia’s public school system.⁶⁶ Gym states that as a teacher, she witnessed “enormous pressure within Philadelphia to have constantly increasing test sources,” and testing created a situation that could be defined as “abusive” to schools and children.⁶⁷ During Philadelphia’s school closure crisis, the administration used these standardized test scores to justify the closing of a school.⁶⁸

The story does not end here, though. The voices of educators, students, and parents are starting to be heard. As Gym described the shutting down of Philadelphia schools, she also spoke about one predominately Black school faced with the option to stay in the Philadelphia school district and receive no funding or be converted to a charter school.⁶⁹ Instead of allowing the Philadelphia School District to push them around, a coalition of teachers, students, and parents formed and overwhelmingly won a vote to stay in Philadelphia’s school district and receive more resources.⁷⁰

In other parts of the country, teachers refuse to administer tests, and students walk out of standardized tests. In Seattle, a group of twenty teachers at Garfield High School organized and refused to administer the Measure of Academic Progress (“MAP”) Test with support of students who refused to take the test.⁷¹ The teachers and students won the day—the Seattle School District conceded that “high schools may opt-out of MAP in 2013-2014.”⁷² From John Dewey to the teachers and students at Garfield High School, the trend to overthrow standardized testing is alive and well. In the next Section, a socio-

⁶⁴ KEVIN G. WELNER, *LEGAL RIGHTS, LOCAL WRONGS: WHEN COMMUNITY CONTROL COLLIDES WITH EDUCATIONAL EQUITY passim* (State Univ. of NY Press 2001).

⁶⁵ *Id.*

⁶⁶ Jesse Hagopian, *From “Shaming and Blaming” to the “Moral Agenda for Our Time”*, in *MORE THAN A SCORE: THE NEW UPRISING AGAINST HIGH-STAKES TESTING* 231 (Jesse Hagopian ed., 2014) [hereinafter *MORE THAN A SCORE*].

⁶⁷ *Id.* at 240.

⁶⁸ *Id.*

⁶⁹ *Id.*

⁷⁰ *Id.* at 242.

⁷¹ Jesse Hagopian, *Preface: The Testocracy versus the Education Spring*, in *MORE THAN A SCORE*, *supra* note 66, at 7.

⁷² *Id.*

legal perspective shows why standardized tests cannot take all variables into account and are invalid as a measure of student aptitude.

B. History of Federal Legislation on Standardized Tests

While educators, students, and parents have pushed back on standardized testing, most of their cries have fallen on the deaf ears of federal legislators. From the first federal legislation on schooling to the most current, legislators have approached education policy as a one-size-fits-all model. Legislators attempting to quantify and qualify schools, teachers, and students started in 1965 with the passage of the Elementary and Secondary Education Act of 1965 (“ESEA”).⁷³

President Lydon B. Johnson launched an ambitious endeavor creating “The Great Society.”⁷⁴ As a counterpart to the War on Vietnam, President Johnson launched a “War on Poverty” as part of his vision for The Great Society.⁷⁵ The ESEA was President Johnson’s attempt to even the educational playing field by providing federal funding to primary and elementary schools.⁷⁶ Since its original passage in 1965, the ESEA has been reauthorized every five years by presidents making slight changes to the law. Yet, with President Bush’s 2001 reauthorization, the NCLBA, the final façade that the federal government was not directly dictating education standards to states was finally stripped away.⁷⁷

One notable aspect of the ESEA is that it focuses on poverty with specific statutory language stating that “payments under this title will be used for programs and projects . . . which are designed to meet the special educational needs of educationally deprived children in school attendance areas having high concentrations of children from low income families.”⁷⁸ In the same section of the statute are the first proverbial federal shots into education heard round the world by requiring:

(5) that effective procedures, including provision for *appropriate objective measurements of educational achievement*, will be adopted for evaluating at least annually the effectiveness of the programs in meeting *the special educational needs of educationally deprived children*;

(6) *that the local educational agency will make an annual report and such other reports to the State educational agency, in such form and*

⁷³ See Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (codified as amended at 20 U.S.C. ch. 70).

⁷⁴ History.com Eds., *Great Society*, HISTORY (Jan. 5, 2022), <https://www.history.com/topics/1960s/great-society> [https://perma.cc/4PJ9-9U4Z].

⁷⁵ Catherine A. Paul, *Elementary and Secondary Education Act of 1965*, VA. COMMONWEALTH UNIV. LIBR. (Jan. 5, 2022), <https://socialwelfare.library.vcu.edu/programs/education/elementary-and-secondary-education-act-of-1965/> [https://perma.cc/W3AB-9ERJ].

⁷⁶ Pub. L. No. 89-10, § 203(a)(1), 79 Stat. at 28.

⁷⁷ See generally Paul, *supra* note 75.

⁷⁸ § 205(a)(1), 79 Stat. at 30.

containing such information, as may be reasonably necessary to enable the State educational agency to perform its duties under this title, *including information relating to the educational achievement of students participating in programs carried out under this title, and will keep such records* and afford such access thereto as the State educational agency may find necessary to assure the correctness and verification of such reports⁷⁹

The next section entitled “Assurances from States” requires that “Any State desiring to participate in the [funding] program of this title Shall submit through its State educational agency to the [federal] commissioner an application,” which includes “the results of objective measurements.”⁸⁰ It is impossible to divorce the tying of federal funding to objective measures to Section 604 of the ESEA stating:

Nothing contained in this Act shall be construed to authorize any department, agency, officer, or employee of the United States to exercise any direction, supervision, or control over the curriculum, program of instruction, administration, or personnel of any educational institution or school system, or over the selection of library resources, textbooks, or other printed or published instructional materials by any educational institution or school system.⁸¹

The mere fact that federal education funding is tied to state objective measurements demonstrates the federal imprimatur on schools, which began with this first federal legislative foray into education law and policy. Aside from tying federal funding to objective measurements—i.e., standardized testing—there is another noteworthy aspect of the ESEA. While this legislation addresses funding for students with disabilities and students who experience poverty—nowhere in the thirty-two pages of the law does it mention that objective measurements must consider and accommodate test takers with disabilities or students experiencing poverty, homelessness, or other extraordinary circumstances.⁸²

One can only speculate if President Johnson and the drafters of the ESEA knew the prohibition against federal interference into state education would soon be discarded. In 1988, President Ronald Reagan signed the Hawkins-Stafford Elementary and Secondary School Improvement Act into law.⁸³ Introduced by Democrat Augustus F. Hawkins, the chairman of the House Education and Labor Committee, and Republican Robert T. Stafford, this bipartisan act required schools and districts to document achievement levels

⁷⁹ § 205(a)(5)-(6), 79 Stat. at 31.

⁸⁰ § 206(a)(3), 79 Stat. at 31-32.

⁸¹ § 604, 79 Stat. at 57.

⁸² *See generally* Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (codified as amended at 20 U.S.C. ch. 70).

⁸³ Paul, *supra* note 75.

via standardized testing of students to receive federal funding—as required by the ESEA.⁸⁴ One small, insidious change in this Act became the cornerstone for the NCLBA; it stated that when students did not show academic progress, Congress would siphon federal funding away.⁸⁵

From president to president, the rigors of testing became greater, but these presidents gave no special consideration to students who experience poverty, homelessness, food insecurity, medical insecurity, household abuse, or neglect. Instead, in 1994, President Bill Clinton heaped on more testing through the Improving America's Schools Act ("IASA").⁸⁶ Unfortunately, President Clinton's IASA did nothing to rid educational funding from the constraints of performance on standardized testing. Instead, it added on math, reading, and language arts standards to assess federal funding to schools.⁸⁷ It was also the first time that the phrase "adequate yearly progress" found its way into federal legislation.⁸⁸

The time became ripe for a reauthorization of the ESEA focused on standardized test scores. In 2001, President Bush reauthorized the ESEA, giving it the now infamous title—No Child Left Behind Act.⁸⁹ The revamped section of the Act required that schools continued to measure and record "adequate yearly progress."⁹⁰ President Johnson's admonition that the federal government should not be overly involved in the state's educational system went out with the bathwater, as the saying goes. President Bush's NCLBA required that state educational agencies make annual reports to the federal government.⁹¹ Schools not making adequate yearly progress face swift punitive measures. If a state fails to meet any of the requirements of this section, other than the requirements described in paragraph (1), then "the Secretary may withhold funds for State administration under this part until the Secretary determines that the State has fulfilled those requirements."⁹²

One substantive change made by President Bush's NCLBA was to begin collecting data in the annual state report on student subject sections, including gender, race, ethnicity, English proficiency status, migrant status, and disability and economic status, specifically for disadvantaged students.⁹³ While

⁸⁴ Educ.Wk. Staff, *Hawkins Leaves Legacy on Accountability*, EDUC.WK. (Nov. 15, 2007), <https://www.edweek.org/education/hawkins-leaves-legacy-on-accountability/2007/11> [https://perma.cc/6D2W-6JXE].

⁸⁵ *See id.*

⁸⁶ *See generally* Improving America's Schools Act (IASA) of 1994, Pub. L. No. 103-382, 108 Stat. 3518 (codified as amended at 20 U.S.C. ch. 70).

⁸⁷ *See* Pub. L. No. 103-382, § 6361(a)(1)–(2); *see also* Paul, *supra* note 75.

⁸⁸ Educ.Wk. Staff, *Every Student Succeeds Act: Adequate Yearly Progress*, EDUC.WK. (Sept. 10, 2004), <https://www.edweek.org/policy-politics/adequate-yearly-progress/2004/09> [https://perma.cc/3WC9-H3T7].

⁸⁹ Paul, *supra* note 75.

⁹⁰ No Child Left Behind Act (NCLB) of 2001, Pub. L. No. 107-110, § 1111(b)(2)(B), 115 Stat. 1425, 1446–47 (2002) (codified as amended in scattered sections of 15 U.S.C., 20 U.S.C., 42 U.S.C. & 47 U.S.C.).

⁹¹ § 1111(h), 115 Stat. at 1457.

⁹² § 1111(g), 115 Stat. at 1457.

⁹³ § 1111(h)(1)(C)(i), 115 Stat. at 1457–58.

President Bush's NCLBA collected this data, it did nothing to address accommodations for students other than students with disabilities and English language learners.⁹⁴ This short overview of the NCLBA is not exhaustive but merely illustrative of the increasing punitive means by which education administrators and legislators use standardized tests without considering other test-taker variables.

The most recent iteration of the reauthorized ESEA came during the Presidency of Barack Obama. President Obama's Every Student Succeeds Act ("ESSA") did a complete 180 degree turn by abolishing federal penalties for struggling schools.⁹⁵ In addition, under the ESSA, states cannot consider only test scores for funding decisions but may consider school quality factors, which include kindergarten readiness, school climate and safety, and chronic absenteeism.⁹⁶ While President Obama's bill was arguably less punitive than President Bush's NCLBA and teachers celebrated the death knell of the NCLBA,⁹⁷ there was still no accounting for the variables mentioned above. Moreover, while the ESSA still required recording of student data based on disaggregated subgroups, it made no accommodations even with all of the data at hand.

A recurrent theme in all of the reauthorized versions of the ESEA is the reliance on the Individuals with Disabilities Education Act to provide accommodations for standardized test takers with disabilities.⁹⁸ The only other group slated into the reauthorized versions have been English Language Learners ("ELL").⁹⁹ There are still no accommodations for students on free-and-reduced lunch, students who are experiencing homelessness, students who are diagnosed with a disability after taking the standardized test, students who are in a period of bereavement, and students experiencing the many other variables that impact their performance during a snapshot, standardized examination.

As the data later in this Article demonstrates, standardized tests taken between President Lyndon Johnson and President Biden's administrations have yet to accommodate all the possible variables present in a student's life. This lack of accommodation for students ultimately raises the question of whether or not standardized testing is truly as standardized, valid, and reliable as corporate test preparation and test creation companies would have one believe.

⁹⁴ The author would like to note that accommodations for students with disabilities and English language learners are sensible in every way. The thrust of the argument is that there are other variables that impact a student's test taking ability that are not accounted for in legislation.

⁹⁵ Jones, *supra* note 40.

⁹⁶ *Id.*

⁹⁷ See, e.g., Kristen Hwang, *Valley Teachers Celebrate Fall of No Child Left Behind*, DESERT SUN (Dec. 10, 2015), <https://www.desertsun.com/story/news/education/2015/12/10/valley-teachers-celebrate-fall-no-child-left-behind/77113036/> [<https://perma.cc/QKZ8-3CNG>].

⁹⁸ See generally Individuals with Disabilities Education Act (IDEA), Pub. L. No. 91-230, 84 Stat. 121 (1970) (codified as amended at 20 U.S.C. ch. 33).

⁹⁹ Individuals with Disabilities Education Act Amendments for 1997, Pub. L. No. 105-17, 111 Stat. 37, 42-43 (codified as amended at 20 U.S.C. ch. 33) (adding definitions to include children with language impairments).

II. AN EXAMINATION OF THE PRELIMINARY SCHOLASTIC APTITUDE TEST

When Lewis Terman and Carl C. Brigham¹⁰⁰ first created the SAT, they may have imagined it only being used in a handful of states.¹⁰¹ Nonetheless, the College Board administers and owns the SAT and a whole new suite of tests, which includes the Preliminary SAT for grades eight and nine (“PSAT 8/9”),¹⁰² are now distributed worldwide.¹⁰³ While the College Board is nominally a not-for-profit, it generated over \$1.2 billion in revenue in 2020¹⁰⁴ while their former CEO, Gaston Caperton, raked in a \$1.3 million salary.¹⁰⁵ As of this writing, there are twelve digital testing dates available to students for the PSAT 8/9.¹⁰⁶ So while students without accommodations in standardized testing languish, test makers have built a global empire. Not to mention the horde of test preparation companies promising higher scores in exchange for cash—an industry itself worth almost one billion dollars.¹⁰⁷

As the PSAT 8/9 is a nationwide test, it will serve as our entryway into statistical analysis. The PSAT 8/9 states that it measures:

[T]he skills and knowledge (at the level of content and skill appropriate for eighth and ninth grades) that research shows are the most important for success in college and career. The Reading Test measures comprehension and reasoning skills, and focuses on close reading of passages in a wide array of subject areas. The Writing and Language Test measures a range of skills, including command of evidence, expression of ideas, and the use of standard English conventions in grammar and punctuation. The Math Test covers a range of math practices, with an emphasis on problem solving, modeling, using tools strategically, and using algebraic structure.¹⁰⁸

¹⁰⁰ Genxing Zhan, *The SAT: the Origin and the Present*, *Educ 300: Education Reform, Past and Present*, TRINITY COLL. (May 6, 2016), <https://commons.trincoll.edu/edreform/2016/05/the-sat-the-origin-and-the-present/> [<https://perma.cc/2TH5-5WFP>].

¹⁰¹ NAT'L EDUC. ASS'N, *supra* note 7.

¹⁰² See generally *SAT Suite of Assessments*, COLL. BD., <https://collegereadiness.collegeboard.org/about> [<https://perma.cc/S6GX-KFU4>].

¹⁰³ *Id.*

¹⁰⁴ *How Much Does the College Board Make Off the SAT and AP Exams?*, FIN. SAMURAI, <https://www.financialsamurai.com/how-much-does-the-college-board-make-off-the-sat-and-ap-exams/> [<https://perma.cc/E3NV-UMF8>].

¹⁰⁵ Tamar Lewin, *Backer of Common Core School Curriculum is Chosen to Lead College Board*, N.Y. TIMES (May 16, 2012), <https://www.nytimes.com/2012/05/16/education/david-coleman-to-lead-college-board.html#:~:text=to%20step%20down.,Mr.,the%20American%20Federation%20of%20Teachers> [<https://perma.cc/75RL-5WYE>].

¹⁰⁶ *Digital Testing – Important Dates*, COLL. BD., <https://digitaltesting.collegeboard.org/important-dates> [<https://perma.cc/ANK3-UExc>].

¹⁰⁷ Susan Taylor, *Examining the Test Prep Industry for Local Business Angles* NAT'L. CTR. FOR BUS. JOURNALISM (July 23, 2019), <https://businessjournalism.org/2019/07/test-prep-industry/> [<https://perma.cc/J7ZH-JR2F>].

¹⁰⁸ COLL. BD., *PSAT 8/9 UNDERSTANDING SCORES 2021-22*, at 3 (2021), <https://collegereadiness.collegeboard.org/pdf/psat-8-9-understanding-scores-april-2019.pdf> [<https://perma.cc/27P2-6XN>].

Noticeably missing from the measurement proffered by the College Board is any mention of material learned through a teacher's curriculum. While standardized tests are: 1) typically summative and 2) specific or general topics may vary across districts and states—content-wise—there is no mention of a connection to the school's actual curriculum.¹⁰⁹ While summative exams, or exams that look at cumulative knowledge, are helpful in the classroom because they are generally tied to a specific unit or theme, there is no such coherence in standardized tests which are divorced from the classroom experience.¹¹⁰ As discussed in Section I.A., this is one of the main complaints lodged against standardized test scores.¹¹¹ Aside from the primary thrust of this Article—arguing that standardized testing is invalid as it cannot account for a plethora of variables—it raises the question, what does standardized testing actually test for, and is it worth taking students out of classrooms for roughly thirty days?¹¹²

Perhaps the most interesting part of the examination process is how the PSAT 8/9 is scored:

Students receive a total score that is the sum of their scores on the two sections (Evidence-Based Reading and Writing and Math). To calculate section scores, we first compute the student's raw score—the number of questions correctly answered—for each section. Nothing is deducted for incorrect answers or for unanswered questions.

Next, we convert each of the raw section scores to a scaled score of 120 to 720. This conversion process adjusts for slight differences in difficulty among versions of the test and provides a score that is consistent across different versions. The scaled scores are the scores provided on score reports.¹¹³

The College Board itself recognizes there is no such thing as a perfectly scaled score.¹¹⁴ While mathematics, reading, and writing are fundamental to a student's learning, innumerable other factors contribute to a student's knowledge, and the ability to express this knowledge cannot always be accounted for on a standardized test.

The College Board states that it “considers all reasonable requests for accommodations needed by students with *documented disabilities*,”¹¹⁵ and

9].

¹⁰⁹ See CHERYL DURWIN & MARLA REESE-WEBER, *EDPSYCH MODULES 535* (Terri Accomazzo, 3rd. ed. 2018).

¹¹⁰ *Formative and Summative Assessments*, YALE UNIV., <https://poorvucenter.yale.edu/Format-ive-Summative-Assessments> [<https://perma.cc/KJM4-RE5P>].

¹¹¹ HUTT & SCHNEIDER, *supra* note 47, at 22.

¹¹² Robelen, *supra* note 57.

¹¹³ COLL. BD., *supra* note 108, at 3.

¹¹⁴ Tamar Lewin, *College Board Corrects Itself on Test Score*, N.Y. TIMES (May 15, 2003), <https://www.nytimes.com/2003/05/15/us/college-board-corrects-itself-on-test-score.html> [<https://perma.cc/W8N8-H3NR>].

¹¹⁵ *Typical Exam Accommodations*, COLL. BD., <https://accommodations.collegeboard.org/typical->

available accommodations are not limited to those listed.¹¹⁶ The Board approves those accommodations¹¹⁷—and others—to ensure that eligible *students with disabilities* face no barriers when taking its exams. While the focus on students with disabilities is by no means something to disregard, it is palpable proof that the College Board has blind spots for students facing poverty, abuse, or food insecurity.

The College Board simply does not offer accommodations for students who face adversity. School districts also falter by giving into the idea of separating the wheat from the chaff as it is easier to let a single, standardized test determine the placement of students. The College Board has attempted to address adversities faced by students but only on their college admissions examination, the SAT.¹¹⁸ One immediate and obvious problem is that by the time students reach the SAT examination—i.e., junior or senior year of high school—schools have already placed the students into lower-level tracks based on the College Board’s suite of other standardized tracking tests.¹¹⁹ In other words, the College Board begins addressing student adversity issues far too late.

In addition, both advocates for and against affirmative action have come out against the adversity index,¹²⁰ making its implementation even more difficult. One affirmative action critic, Manhattan Fellow Heather Mac Donald, stated that “[b]eing raised by a single parent will also be a plus factor. Such a scheme penalizes the bourgeois values that make for individual and community success.”¹²¹ While Mac Donald takes an oblivious view of the impact that community and family have on student success—she is not alone in her criticism. Advocates of affirmative action such as Bart Grachan, Associate Dean for Progress and Completion at LaGuardia Community College, stated that:

Centuries of racial animus and legislation have created and maintained steady segregation of wealth and educational accumulation. Decades of housing policy, practice, and legislation have made sure that there are ‘good’ and ‘bad’ neighborhoods,

accommodations [https://perma.cc/4LBU-N3FB].

¹¹⁶ *Other Accommodations*, COLL. BD., https://accommodations.collegeboard.org/typical-accommodations/other [https://perma.cc/9BGF-GNG8].

¹¹⁷ *How Accommodations Work*, COLL. BD., https://accommodations.collegeboard.org/how-accommodations-work [https://perma.cc/R9Z7-7KCH].

¹¹⁸ Scott Jaschik, *New SAT Score: Adversity*, INSIDE HIGHER ED (May 20, 2019), https://www.insidehighered.com/admissions/article/2019/05/20/college-board-will-add-adversity-score-everyone-taking-sat [https://perma.cc/6UGW-JKK5].

¹¹⁹ HUTT & SCHNEIDER, *supra* note 47, at 7.

¹²⁰ See Scott Jaschik, *Measuring Adversity*, INSIDE HIGHER ED (Feb. 28, 2017), https://www.insidehighered.com/news/2017/02/28/college-board-pilots-new-way-measure-adversity-when-considering-applications-some [https://perma.cc/L3XS-7REW].

¹²¹ Heather Mac Donald, *The College Board Plans to Introduce a New “Adversity Score” As a Backdoor to Racial Quotas in College Admissions*, CITY-J. (May 16, 2019), https://www.city-journal.org/college-boards-sat-adversity-score [https://perma.cc/3PZ2-YBQK].

where tax distribution and local funding separation has ensured the steady maintenance of ‘good’ and ‘bad’ schools, Grachan wrote. Not only does the new system not acknowledge this, he added, the index “is a self-serving attempt to justify their continued existence and contributions to the inequities of the education system.”¹²²

As this overview of the history and educational policies behind the College Board, PSAT, and SAT demonstrate, there are not many fans of the College Board’s suite of tests. While there is an argument that the tests simplify decision-making for course placement and college admission, the data in the following sections proves this claim is patently wrong. Furthermore, the data proves that the PSAT 8/9 is not a reliable tool to make decisions as it does not account for many decisive variables.

A. *The Data Hides: The College Board’s Presentation of PSAT 8/9 Data and the Error in Their Presentation*

Students are required to take the PSAT 8/9 in numerous states, but this Article explores the data of five different states as presented by the College Board. According to the College Board, the PSAT 8/9 was created to help students measure whether or not they are on track for high school, identify the subjects students excel or struggle with, and help students identify potential future careers.¹²³ This test comes at a pivotal time in a student’s educational career as the Education Commission of the States found that “the importance of a smooth transition from 8th grade to 9th grade cannot be emphasized enough, as this transition will determine a student’s success in high school as well as decisions about their post-secondary school life.”¹²⁴ This year is so important that research shows that each course a student fails in eighth grade increases the odds of non-promotion to the ninth to tenth grade by sixteen percent.¹²⁵

To the College Board’s credit, it does include a disclaimer on its data stating that “relationships between test scores and other background or contextual factors are complex and interdependent.”¹²⁶ The states selected were

¹²² Jaschik, *supra* note 118.

¹²³ *What Will Your Child’s Future Look Like?*, COLL. BD. (Sept. 22, 2017), <https://sat-edit.collegeboard.org/pdf/psat-8-9-overview.pdf> [<https://perma.cc/A82H-5VU2>].

¹²⁴ KATHY CHRISTIE & KYLE ZINTH, ENSURING SUCCESSFUL STUDENT TRANSITIONS FROM THE MIDDLE GRADES TO HIGH SCHOOL 1, 2 (Nov. 2008), <https://www.ecs.org/clearinghouse/78/91/7891.pdf> [<https://perma.cc/L3F3-DQZM>].

¹²⁵ Ruth C. Nield & Robert Balfanz, *An Extreme Degree of Difficulty: The Educational Demographics of Urban Neighborhood High Schools*, U. PENN. 1, 14 (Apr. 2006), https://repository.upenn.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1025&context=gse_pubs [<https://perma.cc/6ZJD-SQB2>].

¹²⁶ COLL. BD., SAT SUITE OF ASSESSMENTS ANNUAL REPORT – ILLINOIS 2 (Oct. 25, 2018), <https://reports.collegeboard.org/pdf/2018-illinois-sat-suite-assessments-annual-report.pdf> [<https://perma.cc/GLJ9-6Q83>].

drawn at random and make up a representative sample of the United States. Below is the data provided by the College Board broken down into states' potential scores, percentage receiving those scores by grade level, and the mean score of each grade level for the particular state.

Table 1: PSAT 8/9 Scores in 2018

State	Number of Students Taking Test by Grade	Score Ranges	Grade 9 Scores by Percent	Grade 8 Scores by Percent	Mean of Grade 9	Mean of Grade 8
Ill. ¹²⁷	9th: 124,177	1400-1440:	0%	0%	880	835
		1200-1390:	5%	2%		
		1000-1190:	21%	13%		
	8th: 52,709	800-990:	40%	42%		
		600-790:	32%	40%		
240-590:	3%	4%				
Ind. ¹²⁸	9th: 16,908	1400-1440:	0%	0%	891	834
		1200-1390:	4%	1%		
		1000-1190:	23%	13%		
	8th: 8,539	800-990:	43%	43%		
		600-790:	28%	40%		
240-590:	2%	3%				
Mass. ¹²⁹	9th: 10,501	1400-1440:	0%	0%	872	847
		1200-1390:	3%	1%		
		1000-1190:	21%	14%		
	8th: 2,645	800-990:	42%	45%		
		600-790:	30%	36%		
240-590:	4%	3%				
N.Y. ¹³⁰	9th: 17,862	1400-1440:	0%	0%	858	859
		1200-1390:	2%	2%		
		1000-1190:	19%	19%		
	8th: 3,893	800-990:	41%	41%		
		600-790:	34%	34%		
240-590:	4%	4%				

¹²⁷ *Id.* at 16–17.

¹²⁸ COLL. BD., SAT SUITE OF ASSESSMENTS ANNUAL REPORT – INDIANA 16–17 (Oct. 25, 2018), <https://reports.collegeboard.org/pdf/2018-indiana-sat-suite-assessments-annual-report.pdf> [https://perma.cc/M738-DUUG].

¹²⁹ COLL. BD., SAT SUITE OF ASSESSMENTS ANNUAL REPORT – MASSACHUSETTS 16–17 (Oct. 25, 2018), <https://reports.collegeboard.org/pdf/2018-massachusetts-sat-suite-assessments-annual-report.pdf> [https://perma.cc/4TV8-HXCU].

¹³⁰ COLL. BD., SAT SUITE OF ASSESSMENTS ANNUAL REPORT – NEW YORK 16–17 (Oct. 25, 2018), <https://reports.collegeboard.org/pdf/2018-new-york-sat-suite-assessments-annual-report.pdf> [https://perma.cc/BV5K-FRAG].

Tex. ¹³¹	9th:	1400-1440:	0%	0%	850	800
	112,122	1200-1390:	3%	2%		
		1000-1190:	17%	12%		
	8th:	800-990:	39%	31%		
	171,520	600-790:	36%	47%		
		240-590:	4%	7%		

As one can see under the score ranges column, the relative frequency¹³² is off in two ways when looking to Illinois as a sample state for illustrative purposes.¹³³ First, the score range is consistent in increments of 200 from 1400 down to 600 until the chart becomes inconsistent, skipping 360 points from 600 down to 240.¹³⁴ This complicates any statistical analysis as the relative frequency, or the distribution of scores is now skewed—to create a perfect bell curve, perhaps—and it is impossible to tell whether the three percent of test-takers in that range scored near 590 or 240.¹³⁵ In a normal distribution of ranges, the ranges would descend in a standard order, i.e., going down by a uniform range of 200 points—for example, 1000–800, 800–600, and so on. Second, the class width is not uniform between the scores themselves. The highest score range, 1400–1440, is only a forty-point range, whereas the lowest score range, 250–590, is a 340-point range, which again makes comparison impossible as the width should also be consistent to do any meaningful statistical analysis.¹³⁶ In other words, the College Board’s presentation of the data is wrong *ab initio*. This is the same issue for all the states in the table and rather display the statistical calculations for each; suffice it to say, the data is, in essence, almost worthless. It also raises the question of whether an organization that employs professional test takers and likely mathematicians would not know this error in class widths would make statistical analysis much more difficult.

To reconcile the data as best as possible, it is imperative that we create class widths that make sense. Therefore, we should now base the data set on a class range of 200—in essence, creating 200–400, 400–600, and so on.¹³⁷ Following this new class range, we should distribute three percent of Illinois ninth-grade students: one percent in the 200–400 class width and two percent in the 400–600 class width.¹³⁸ While the College Board gives the mean for

¹³¹ COLL. BD., SAT SUITE OF ASSESSMENTS ANNUAL REPORT – TEXAS 16–17 (Oct. 25, 2018), <https://reports.collegeboard.org/pdf/2018-texas-sat-suite-assessments-annual-report.pdf> [<https://perma.cc/V4PZ-BPW6>].

¹³² *Bitesize Maths*, BRIT. BROAD. CHANNEL (Jan. 6, 2022), <https://www.bbc.co.uk/bitesize/guides/zjw7hyc/revision/1> [<https://perma.cc/V5FD-BDLH>].

¹³³ Illinois will be used as an illustrative example because the author is most familiar with Illinois and all five states hold relatively large school districts.

¹³⁴ COLL. BD., *supra* note 126, at 16–17.

¹³⁵ *Id.*

¹³⁶ *Id.*

¹³⁷ *Id.*

¹³⁸ *Id.* It should be noted that while this data set is not exact, it is closest to what the College Board gives to consumers based on the principles of statistical analysis.

Illinois ninth-grade students, it does not provide the standard deviation or a means to equate the standard deviation.¹³⁹ Table 2 provides the normal distribution equation below, where μ standards for the mean and σ^2 for the standard deviation.¹⁴⁰

Table 2: Normal Distribution Formula

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$

μ = Mean

σ = Standard Deviation

$\pi \approx 3.14159 \dots$

$e \approx 2.71828 \dots$

In addition, the College Board only gives three distributions—the 75th, 50th, and 25th percentiles—and omits the maximum and minimum distributions making statistical analysis a guessing game.¹⁴¹ With such a large data set as 124,177 Illinois ninth-graders, one sensible standard deviation is 200. Therefore, using the normal distribution formula of 880¹⁴² as the mean and 200 as a standard deviation, the approximation comes out as described in Table 3.

¹³⁹ *Id.*

¹⁴⁰ *Continuous Probability Distributions*, DUKE UNIV. (May 13, 2021), <https://sites.nicholas.duke.edu/statsreview/continuous-probability-distributions/> [<https://perma.cc/CBA7-NTGW>].

¹⁴¹ COLL. BD., *supra* note 126, at 16–17.

¹⁴² *Id.*

Table 3: College Board Normal Distribution Curve¹⁴³

There are many issues with this data set. First, the chart exemplifies the expertise of test makers hiding data in plain sight—as the task for creating an assessment instruct that “with a handful of items, yields valid norm-referenced interpretations of a student’s status regarding a substantial chunk of content.”¹⁴⁴ The methodology of achieving this means is that test-takers develop items that are answered correctly by roughly half the students: “developers avoid items that are answered correctly by too many or by too few students.”¹⁴⁵ In other words, a valid standardized test would create a normal distribution because approximately half of the students would get questions right, and half would get questions wrong. Unfortunately, the College Board’s presentation of data neither allows for that assumption to be tested nor seems to bear out a normal distribution.¹⁴⁶ In sum, the College Board’s data set questions the validity of the PSAT 8/9.

Second, standardized testing does not consider several variables discussed throughout this article, such as poverty, homelessness, and food insecurity. Finally, these charts are deceptive as there is the issue of confounded causation.¹⁴⁷ The term “confounded causation” refers to three factors that test takers and state and federal administrators use to evaluate educational quality and progress; in reality, these factors may be correlative at best.¹⁴⁸ These factors include 1) the material taught in the classroom, 2) a student’s innate

¹⁴³ *Id.* This is the normal distribution after adjusting for the standard deviation, which is mere guess work.

¹⁴⁴ W. James Popham, *Why Standardized Tests Don’t Measure Educational Quality*, ASS’N FOR SUPERVISION & CURRICULUM DEV. (1999), <http://www.ascd.org/publications/educational-leadership/mar99/vol56/num06/Why-Standardized-Tests-Don%27t-Measure-Educational-Quality.aspx> [<https://perma.cc/GV9B-2WW4>].

¹⁴⁵ *Id.*

¹⁴⁶ COLL. BD., *supra* note 126, at 16–17.

¹⁴⁷ Popham, *supra* note 144.

¹⁴⁸ *Id.*

intellectual ability, and 3) the student's out-of-school experiences.¹⁴⁹

While the College Board states that it scales the PSAT 9/8 scores to ensure comparable results across different test forms, it stretches credulity to believe the College Board can convert raw scores onto a consistent and standardized scale when the test is flawed from the inception.¹⁵⁰ Similar to legislation surrounding standardized test scores and data collection, the College Board, in June 2015, aligned its collection of race and ethnicity categories with the U.S. Department of Education.¹⁵¹ While collecting these categories can help break down achievement gaps, it does nothing except paint a picture of the systemic racism in testing already proven to exist.¹⁵² Instead of addressing the root cause of test-taking differences, such as the race disparity in who can afford test preparation courses, the College Board merely collects data on the issue.¹⁵³

Educational researcher Sigal Alon found “the black-white gap is especially pronounced in the use of high school courses, private courses, and private tutors [for white students].”¹⁵⁴ With private courses, tutors, and test preparation also come issues of wealth. In a 2015 analysis by researchers at the National Center for Fair and Open Testing, students with a family income of less than \$20,000 scored lowest on standardized tests.¹⁵⁵ The linkage between test preparation and test outcomes is undeniable as even the College Board admits that there are benefits associated with test preparation.¹⁵⁶

While the College Board's presentation of data is, at first glance, compelling, there is a clarion call across the nation against the PSAT 8/9, the SAT, and the entirety of the extensive SAT suite the College Board has to offer to schools and states. For example, in a lawsuit filed on behalf of the Compton Unified School System against the Regents of the University of California System, the district claims that the University of California system is violating state civil rights laws by requiring applicants to take the SAT or its counterpart, the ACT.¹⁵⁷ In addition, the lawsuit alleges that the standardized

¹⁴⁹ *Id.*

¹⁵⁰ COLL. BD., *supra* note 126, at 16–17.

¹⁵¹ *Collection and Reporting of Student Race and Ethnicity Data*, COLL. BD., <https://research.collegeboard.org/about-us/changes-to-race-ethnicity-reporting> [<https://perma.cc/E4TF-AWDP>].

¹⁵² John Rosales & Tim Walker, *The Racist Beginnings of Standardized Testing*, NAT'L EDUC. ASS'N. (Mar. 20, 2021), <https://www.nea.org/advocating-for-change/new-from-nea/racist-beginnings-standardized-testing> [<https://perma.cc/N2R6-QSW7>].

¹⁵³ COLL. BD., *supra* note 126, at 16–17; *see also* Sigal Alon, *Racial Differences in Test Preparation Strategies: A Commentary on Shadow Education, American Style: Test Preparation, the SAT and College Enrolment*, 89 SPECIAL FORCES 463, 470 (2010).

¹⁵⁴ Alon, *supra* note 153, at 463.

¹⁵⁵ Kristina Garcia, *SAT and ACT exams undermine low-income students*, DAILY TITAN (Apr. 8, 2021), https://dailytitan.com/opinion/sat-and-act-exams-undermine-low-income-students/article_ec473934-97d4-11eb-a123-2ff94de71f68.html [<https://perma.cc/Q8FN-7Z5Q>].

¹⁵⁶ *See SAT Practice*, COLL. BD., <https://collegereadiness.collegeboard.org/sat/practice> [<https://perma.cc/RP4S-63WS>]; *see also* Alon, *supra* note 153, at 463.

¹⁵⁷ Nanette Asimov, *Students of Color to UC: Drop the SAT or We'll See You in Court*, S.F. CHRON. (Oct. 30, 2019), <https://www.sfchronicle.com/education/article/Students-of-color-to-UC-Drop-the-SAT-or-we-ll-14572455.php> [<https://perma.cc/C75J-H2ZM>].

tests “unlawfully discriminate against disabled, low-income, multilingual and underrepresented minority students.”¹⁵⁸

In a sister case to the Compton Unified School District, *Smith v. Regents of the University of California*,¹⁵⁹ Judge Brad Seligman of the Superior Court of California issued a preliminary injunction in ongoing litigation that requires the University of California system to immediately stop using the SAT and ACT as part of the admissions process.¹⁶⁰ This is one of a myriad of lawsuits filed against the College Board, including a class-action lawsuit regarding the PSAT 8/9 in the Northern District of Illinois.¹⁶¹ The lawsuit, *S. v. College Board*,¹⁶² claims that the College Board sells personally identifiable information collected from test takers on the PSAT 8/9, which may come as no surprise to one who follows the College Board as this complaint has been levied against them numerous times.¹⁶³

While all these issues raise valid concerns, there is nothing more telling than the data—which is patently attempting to hide some flaw. In Part B, the Article discusses social science research’s findings that standardized tests fail to account for extraneous variables. In a general way, it also discusses the impact of these extraneous variables on standardized tests.

B. The Data Fails: How the Existence of Extraneous Variables Invalidates the PSAT 8/9 and Other College Board Tests

It comes as no surprise that factors such as race, gender, socioeconomic status, ELL status, school discipline, and students with disabilities all affect standardized test scores. The question becomes what does the data reflect about the variability of test scores if a student falls into one or more of these categories. This Section will first explore the PSAT 8/9 and variabilities then discuss findings on the SAT.

As the previous section noted, there is no possible way to make a meaningful distribution curve for the data provided by the College Board. However, there is enough data to make conclusions about the PSAT 8/9 and the variables it provides in the discrete data sets, including race and ethnicity, sex, and first language learned.¹⁶⁴ Therefore, using the information from Illinois’s test-takers, we find the following:

¹⁵⁸ Maxwell Strachan, *California Students Are Suing the UC System for Requiring the SAT and ACT*, VICE (Dec. 10, 2019), <https://www.vice.com/en/article/vb555a/california-students-are-suing-the-uc-system-for-requiring-the-sat-and-act> [<https://perma.cc/JSP9-WBK4>].

¹⁵⁹ *Smith v. Regents of the Univ. of Cal.*, No. RG19046222 (Super. Ct. Alameda Cnty. Aug. 31, 2020).

¹⁶⁰ Order Denying Stay of Preliminary Injunction at 1, *Smith v. California*, No. RG19046222 (Super. Ct. Alameda Cnty. Sept. 15, 2020), <https://publiccounsel.org/wp-content/uploads/2022/02/2020.09.15-Order-Denying-Ex-Parte-Application.pdf> [<https://perma.cc/X8EW-H9W5>].

¹⁶¹ Class Action Complaint at 1, *S. v. Coll. Bd.*, 1:19-cv-08068 (N.D. Ill. Dec. 10, 2019).

¹⁶² *Id.*

¹⁶³ *Id.* at 3.

¹⁶⁴ COLL. BD., *supra* note 126, at 9–10.

Table 4: Race/Ethnicity¹⁶⁵

Race/Ethnicity	Test Takers		Mean Score		
	Number	Percent	Total	ERW	Math
Total					
American Indian/Alaska Native	2,169	1%	771	387	384
Asian	8,388	5%	992	489	503
Black/African American	22,052	12%	773	387	386
Hispanic/Latino	43,967	25%	819	408	411
Native Hawaiian/Other Pacific Islander	160	0%	805	405	400
White	71,236	40%	917	463	454
Two or More Races	5,297	3%	892	451	442
No Response	25,640	14%	844	424	419

Table 5: Sex¹⁶⁶

Sex	Test Takers		Mean Score		
	Number	Percent	Total	ERW	Math
Female	87,182	49%	875	444	431
Male	90,211	50%	858	426	431
No Response	1,516	1%	779	394	385

Table 6: First Language Learned¹⁶⁷

First Language Learned	Test Takers		Mean Score			Met Benchmarks ¹			
	Number	Percent	Total	ERW	Math	Both	ERW	Math	None
8th Grade									
English Only	24,757	47%	853	433	420	44%	70%	47%	27%
English and Another Language	7,992	15%	816	407	408	34%	55%	39%	39%
Another Language	3,135	6%	791	392	399	29%	47%	34%	48%
No Response	16,825	32%	825	418	407	38%	60%	42%	35%

When examining Table 4's race and ethnicity data, one can detect socioeconomic factors' impact on students, resulting in obvious patterns emerging, such as American Indian/Alaska Native and Black/African American students scoring in the bottom two places, respectively.¹⁶⁸ In contrast, White and Asian students scored in the top two spots.¹⁶⁹ The range of scores—from American Indian/Alaska Native to Asian—is 281 points, but the most telling statistic is the mean score, which works out to 844.125.¹⁷⁰ The averages of the minority populations fall below this mean: American Indian/Alaska Native, Black/African American, Native Hawaiian or Other

¹⁶⁵ *Id.* at 9 (stating "ERW" means Evidenced-based Reading and Writing).

¹⁶⁶ *Id.*

¹⁶⁷ *Id.* at 10.

¹⁶⁸ *Id.* at 9.

¹⁶⁹ *Id.*

¹⁷⁰ *Id.*

Pacific Islander, and Hispanic/Latino.¹⁷¹ While the no response category introduces an unknown variable, it is patently clear that Asian and White students dominate standardized testing over practically every other racial group except “Two or More Races” and “No Response.”¹⁷²

Turning to sex, female students tend to score higher than male students overall.¹⁷³ However, when looking ahead at female participation in areas such as science, technology, engineering, and mathematics (“STEM”), there are disheartening data points. Throughout a woman’s educational career—from standardized testing in fourth, eighth, and twelfth grades—where they perform equally with their male counterparts, schools still assign more boys to advanced STEM courses than their female counterparts.¹⁷⁴ According to the National Center for Education Statistics, women still earn less than twenty percent of bachelor’s degrees in engineering and computer science, and information science.¹⁷⁵

When it comes to language, it may not be surprising that individuals who speak another language score sixty-two points lower on the PSAT 8/9 than students who only speak English.¹⁷⁶ But, again, we see a hierarchy of scores with English-only speakers coming in the highest, no response coming in second, and bilingual or monolingual students coming in last place.¹⁷⁷ While the data provided by the College Board on the PSAT 8/9 is scarce, this Article looks at aspects of another one of its suite of tests, the SAT, to draw more comparisons based on income, race, and gender. The data presented below corroborates the PSAT 8/9 data, signaling that students who have one indicator¹⁷⁸ against them are at a substantial disadvantage when it comes to standardized tests.

For example, the Oak Park-River Forest School District in Illinois, lauded for its racial integration efforts,¹⁷⁹ still used students’ performances on the SAT, with its results demonstrating massive racial disparities in SAT outcomes. For instance, in 2018-19, in the English and Language Arts achievement portion of the SAT, only a little over twenty-five percent of Black Oak Park students met or exceeded expectations.¹⁸⁰ In comparison, a full

¹⁷¹ *Id.*

¹⁷² *Id.*

¹⁷³ *Id.*

¹⁷⁴ *The State of Girls and Women in STEM*, NAT’L GIRLS COLLABORATIVE PROJECT (March 2022), https://ngcproject.org/sites/default/files/2022-02/ngcp_stateofgirlsandwomeninstem_2022_0.pdf [<https://perma.cc/Z3WJ-GC3Z>].

¹⁷⁵ Matthew McLaughlin, *Can Title IX Change STEM Culture?*, NAT’L SOC’Y OF PROF. ENGINEERS (May/Jun. 2017), <https://www.nspe.org/resources/pe-magazine/may-2017/can-title-ix-change-stem-culture> [<https://perma.cc/BNV7-NDQG>].

¹⁷⁶ COLL. BD., *supra* note 126, at 10.

¹⁷⁷ *Id.*

¹⁷⁸ Jaschik, *supra* note 118.

¹⁷⁹ John Gearen Jr., *Looking Back at a Pivotal Time in Oak Park History*, WEDNESDAY J. OF OAK PARK & RIVER FOREST (Feb. 10, 2021), <https://www.oakpark.com/News/Articles/5-29-2018/Looking-back-at-a-pivotal-time-in-Oak-Park-history/> [<https://perma.cc/M5DU-MCV2>].

¹⁸⁰ *Oak Park & River Forest High School (9-12)*, IL. STATE BD. OF EDUC., <https://www.illinoisreportcard.com/school.aspx?source=trends&source2=sat.details&Schoolid=060162000130001>

eighty-one percent of their white peers met or exceeded expectations.¹⁸¹ For non-low-income students, a full seventy-three percent of students met or exceeded expectations, while only twenty-seven percent of their low-income peers met or exceeded expectations.¹⁸² While various factors may account for these disparate outcomes, the fact that the achievement gap between Black and white students was nearly fifty-six percent and between low-income and non-low-income students was forty-six percent cannot be explained away easily.¹⁸³

The same data plays out at another school district in Illinois, Adlai E. Stevenson High School District 125. Adlai E. Stevenson High School, where per-pupil expenditure is nearly \$8,000 over the state average of approximately \$15,000 per student, data numbers are distributed the same as Oak Park-River Forest's numbers.¹⁸⁴ At Stevenson, thirty-seven percent of Black students are below the performance rates, while only twenty-three percent of white students are at the same level in English and Language Arts.¹⁸⁵ For Mathematics, this number drops for Black students to forty-two percent below the median performance rates but only dips to twenty-one percent for white students.¹⁸⁶ For low-income students, the numbers are even worse. Only thirty-one percent of low-income students met Mathematics expectations, while eighty-four percent of non-low-income students met or exceeded expectations.¹⁸⁷ For English and Language Arts, thirty-eight percent of low-income students met expectations, while a full eighty percent of non-low-income students did the same.¹⁸⁸ Even at a school as well-resourced as Stevenson, the achievement gap based on race and socioeconomic status cannot be accounted for by the woes of low-income students.

When looking at the national research regarding gender implications in education, Peggy Orenstein finds that girls start to fall behind boys by high school because a loss in confidence precedes a drop in academic achievement.¹⁸⁹ Compounding this problem is, as Orenstein finds, the fact that the "gap in standardized test scores between the affluent and the poor is far

[<https://perma.cc/VZB7-7KLC>].

¹⁸¹ *Id.*

¹⁸² *Id.*

¹⁸³ *Id.* The evidence in Orenstein's study demonstrated that teenage girls experience reduced academic expectations from teachers as well as their female peers. PEGGY ORENSTEIN, *SCHOOLGIRLS: YOUNG WOMEN, SELF-ESTEEM, AND THE CONFIDENCE GAP* xx (Anchor Books 2000) (1994). According to Orenstein, this precipitates a lack of confidence in academic ability when compared to the academic confidence of boys their age. *Id.* at xxi.

¹⁸⁴ *Adlai E. Stevenson High School*, IL. STATE BD. OF EDUC., <https://www.illinoisreportcard.com/School.aspx?schoolId=340491250130001> [<https://perma.cc/3W4V-WMGK>] (in 2020, the average spending per student in this district was \$22,464). *But see Illinois Report Card*, IL. STATE BD. OF EDUC., <https://www.illinoisreportcard.com/State.aspx?source=environment&source2=perstudentspending&Stateid=IL> [<https://perma.cc/Y3ZV-QPP8>] (in 2020, the average operational spending per student in Illinois was \$14,747).

¹⁸⁵ *Adlai E. Stevenson High School*, *supra* note 184.

¹⁸⁶ *Id.*

¹⁸⁷ *Id.*

¹⁸⁸ *Id.*

¹⁸⁹ *See* ORENSTEIN, *supra* note 183, at xxi.

greater than any gender gap between low-income boys and girls.”¹⁹⁰ Therefore, Black girls who disengage from the curriculum as a means of resistance and are also in the low-income category are the most severely disadvantaged youth in our education system.

Also, while the Illinois State Board of Education offers data based on female and male performance,¹⁹¹ it does not include gender-specific data in this section because it does not break down male and female performance by race and socioeconomic status. This omission makes a worthwhile comparison difficult, if not impossible. However, when looking at male and female achievement, the data shows that the gap is not as large as the ones solely based on race or socioeconomic status. Thus, this data is skewed by the unavailability of a breakdown by gender based on these two factors.

Turning to the impact of school discipline, one should consider that the effect of school discipline is never benign. For example, school suspensions and expulsions have been associated with a reduced likelihood of high school graduation, decreased chance of post-secondary enrollment,¹⁹² and a higher risk for entry into the juvenile and criminal justice system.¹⁹³ In addition, disciplinary policies disproportionately impact particular subgroups of students, particularly Black students and students with disabilities,¹⁹⁴ as well as Hispanic and Latino students in middle and high school, Native American students with disabilities, lesbian, gay, bisexual, and transgender (“LGBT”) students, and students who are gender non-conforming. Furthermore, when looking to proficiency gaps in standardized tests, researchers have found that:

Students with 2-to-4 referrals were at 2.7 greater odds of scoring below proficiency, whereas students with 5-or-more were at about 4.6 greater odds.¹⁹⁵

Looking at proficiency gaps [in standardized tests], among

¹⁹⁰ *Id.* at 144.

¹⁹¹ See generally *Illinois Report Card*, *supra* note 184.

¹⁹² Robert Balfanz, Vaughan Byrnes, & Joanna Fox, *Sent Home and Put Off Track: The Antecedents, Disproportionalities, and Consequences of Being Suspended in the 9th Grade*, 5 J. APPLIED RSCH. ON CHILD. 2 (2014), <https://digitalcommons.library.tmc.edu/cgi/viewcontent.cgi?article=1217&context=childrenatrisk> [<https://perma.cc/EW9K-54KK>].

¹⁹³ TRACY L. SHOLLENBERGER, RACIAL DISPARITIES IN SCHOOL SUSPENSION AND SUBSEQUENT OUTCOMES: EVIDENCE FROM THE NATIONAL LONGITUDINAL SURVEY OF YOUTH 1997 7 (May 27, 2013), https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/state-reports/racial-disparities-in-school-suspension-and-subsequent-outcomes-evidence-from-the-national-longitudinal-survey-of-youth-1997/Shollenberger_racial-disparities-suspension-draft.pdf [<https://perma.cc/JYP5-6KN5>]; Daniel J. Losen, *Conclusion*, in CLOSING THE SCHOOL DISCIPLINE GAP: EQUITABLE REMEDIES FOR EXCESSIVE EXCLUSION 241 (Daniel J. Losen ed., 2015) [hereinafter CLOSING THE SCHOOL DISCIPLINE GAP].

¹⁹⁴ SHOLLENBERGER, *supra* note 193, at 11; Losen, *Introduction*, in CLOSING THE SCHOOL DISCIPLINE GAP, *supra* note 193, at 2–4, 7–8.

¹⁹⁵ ANDY WHISMAN & PATRICIA CAHAPE HAMMER, THE ASSOCIATION BETWEEN SCHOOL DISCIPLINE AND MATHEMATICS PERFORMANCE: A CASE FOR POSITIVE DISCIPLINE APPROACHES 7 (2014), <https://files.eric.ed.gov/fulltext/ED569903.pdf> [<https://perma.cc/KL6L-7J99>].

students with a single discipline referral there was a 25 percentage point proficiency gap, which increased by about 20 percentage points with each level of increased disciplinary involvement, i.e., 2-to-4 discipline referrals or 5-or-more discipline referrals.¹⁹⁶

This proficiency gap compounds other variables such as race, socioeconomic status,¹⁹⁷ and gender.¹⁹⁸ This section demonstrates all data points to extraneous variables invalidating the PSAT 8/9 and the SAT. Without a methodology for accounting for these variables, which this author doubts one could create, standardized tests are simply anything but standard. The College Board can use an adversity index to try to level the playing field, but the simple matter is there has never been a level playing field. The College Board, state boards of education, and school districts simply cannot overlook that standardized tests cannot and will not ever account for extraneous variables. The next Section will address several solutions to this issue.

III. ABOLITIONIST VISIONS OF THE FUTURE

A. *The Vision for the Future: How Educators Imagine the Future of Testing*

President Reagan quickly disregarded President Johnson's ESEA admonition for the federal government to stay out of state education, and the education reform movement followed swiftly on Reagan's coattails.¹⁹⁹ Policymakers, business leaders, school boards, and even some parents led this education reform movement because they wanted proof their investment in schools was paying dividends.²⁰⁰ Since the education reform movement began, there has been pushback against standardized testing.²⁰¹

Educators have offered alternatives to the rote, standardized forms that students fill out multiple times a school year. Instead, educational researchers Doug Archbald and Fred Newmann looked to undergraduate and graduate program assessments.²⁰² According to Archbald and Newmann, essays and oral examinations are required in most graduate programs and are often considered the most rigorous and valid test of academic competence.²⁰³ In fact, "[t]hey are

¹⁹⁶ *Id.* at 10.

¹⁹⁷ *Id.*

¹⁹⁸ *Id.*

¹⁹⁹ See Valerie Strauss, 'A Nation at Risk' Demanded Education Reform 35 Years Ago. Here's How It's Been Bungled Ever Since., WASH. POST (Apr. 2018), <https://www.washingtonpost.com/news/answer-sheet/wp/2018/04/26/the-landmark-a-nation-at-risk-called-for-education-reform-35-years-ago-heres-how-it-was-bungled/> [https://perma.cc/J5SH-LTFK].

²⁰⁰ DOUG A. ARCHBALD & FRED M. NEWMANN, BEYOND STANDARDIZED TESTING: ASSESSING AUTHENTIC ACHIEVEMENT IN SECONDARY SCHOOL 7 (1988), <https://files.eric.ed.gov/fulltext/E D301587.pdf> [https://perma.cc/5J84-C2HF].

²⁰¹ *Id.*

²⁰² *Id.* at 22.

²⁰³ *Id.*

recognized internationally for the depth with which they assess mastery of specific subjects as well as analytical skills. Usually, they require the student to integrate knowledge and speak extensively.²⁰⁴

While noting that this is method is perhaps the best way to approach student assessment, Archbald and Newmann note schools often have time constraints based on the number of students and allotted time per period.²⁰⁵ Possible solutions to these issues include assessing fewer students at a time and staging the assessments throughout the course of study.²⁰⁶ One appealing factor of this type of testing is that individual teachers can adjust their assessment based on their personal knowledge of the student and the student's progression throughout the school year. However, one issue is educator bias based on gender²⁰⁷ or race.²⁰⁸ While schools and other organizations have attempted to reduce racial bias, there is still the extraneous variable of the teacher's own bias and the student's ability to perform written or verbal activities.²⁰⁹

Another approach proposed by researchers is known as the stealth assessment.²¹⁰ According to former principal research scientist turned professor at Florida State University, Valerie Shute, stealth assessments passively collect data on students' knowledge over a semester, year, or entire school career.²¹¹ A school's assessment of students would take place behind the scenes, using software such as Khan Academy, where students can log on daily, learn, and then answer questions.²¹² This process would eliminate the loss of instructional time, reduce cost, and temper test anxiety.²¹³ While there are logistical issues for rolling out a nationwide program of passive testing, such as access to computers, software, and professional training, this system would allow teachers to recover crucial instructional time with students.²¹⁴

Other novel approaches have started using performance exams instead of standardized tests.²¹⁵ For example, a pilot Milwaukee Public School program

²⁰⁴ *Id.*

²⁰⁵ *Id.*

²⁰⁶ *Id.* at 24.

²⁰⁷ See generally Myra Sadker & David Sadker, *Sexism in the Classroom: From Grade School to Graduate School*, 67 PHI DELTA KAPPAN 512, 512–15 (Mar. 1986).

²⁰⁸ See generally Madeline Will, *Teachers Are as Racially Biased as Everybody Else, Study Shows*, EDUC.WK. (June 9, 2020), <https://www.edweek.org/teaching-learning/teachers-are-as-racially-biased-as-everybody-else-study-shows/2020/06> [<https://perma.cc/5LX2-8BAE>].

²⁰⁹ See generally Jordan G. Starck, Travis Riddle, Stacey Sinclair & Natasha Warikoo, *Teachers Are People Too: Examining the Racial Bias of Teachers Compared to Other American Adults*, 49 EDUC. RESEARCHER 273 (May 2020).

²¹⁰ Jordan Walker, *How to Measure Student Progress Without Standardized Tests*, OTUS, <https://otus.com/measure-student-progress-without-standardized-testing/> [<https://perma.cc/TFD7-H2AZ>].

²¹¹ Anya Kamenetz, *What Schools Could Use Instead of Standardized Tests*, NAT'L PUB. RADIO (Jan. 6, 2015), <https://www.npr.org/sections/ed/2015/01/06/371659141/what-schools-could-use-instead-of-standardized-tests> [<https://perma.cc/S2NP-BS9D>].

²¹² *Id.*

²¹³ *Id.*

²¹⁴ *Id.*

²¹⁵ See generally Bob Peterson & Monty Neill, *Alternatives to Standardized Tests*, 13

assesses fourth and fifth graders by having them perform a three to five-minute oral presentation, while students in fourth, fifth, eighth, eleventh, and twelfth grades write and revise an essay over two days.²¹⁶ These essays are then judged anonymously and randomly by two teachers using a scale of one to four, with the final score based on the sum of the two readers to eliminate as much bias as possible.²¹⁷ However, if there is more than one point difference between the readers, a third reader is brought in to score the paper.²¹⁸ There are certainly downsides to this method, including the considerable time it takes to execute and the temptation to teach to the test. Teaching to the test is the idea that students will have unique ways to express their knowledge, and the reduction of subjectivity—that standardized tests claim to eliminate—makes the performance examination model a promising instrument for the future.²¹⁹

A final, novel approach is the collection of social and emotional skills surveys.²²⁰ As research has found, “at least half of long-term chances of [student] success are determined by nonacademic qualities like grit, perseverance, and curiosity.”²²¹ One school district piloting this type of program is the Montgomery County Public School district in Maryland which surveys teachers and students on social and emotional factors to guide internal decision-making.²²² The idea behind these surveys is to gauge student engagement in schools, which promising studies have shown do a better job of predicting persistence and grade point average than the SAT.²²³ While schools may be reluctant to ditch traditional, standardized testing, student engagement is an important measure regardless of whether or not students still must take standardized tests.

As this brief overview of the literature demonstrates, there are many educators voicing alternatives to high-stakes, standardized tests. These alternative suggestions are being made despite the proponents of the status quo highlighting the risk school districts are taking when implementing new assessment methods and the belief that standardized tests are really objective. As the sections above have demonstrated, though there is no such thing as an objective examination—even the review of this literature reveals problems in alternative forms of assessment. Despite these potential problems, it is time to abolish the standardized testing regime that propagates and perpetuates racial, gender, economic, and other disparities in higher education and career trajectory choices.

RETHINKING SCHS. (Spring 1999), <https://rethinkingschools.org/articles/alternatives-to-standardized-tests/> [<https://perma.cc/Y5BL-UMGG>].

²¹⁶ *Id.*

²¹⁷ *Id.*

²¹⁸ *Id.*

²¹⁹ *Id.*

²²⁰ Kamenetz, *supra* note 211.

²²¹ *Id.*

²²² *Id.*

²²³ *Id.*

B. A Future without Standardized Tests: A Suggestion for Moving Forward

It is helpful first to revisit what researchers consider a standardized test. A standardized test is distinguished from other assessments by two qualities: 1) they are created by testing experts at test publishing companies; and 2) a trained examiner gives all students the test under the same conditions.²²⁴ There may be formative or summative assessments of a student's learning and knowledge in a classroom.²²⁵ An assessment includes "any and all procedures used to collect information and make inferences or judgments about an individual or a program."²²⁶ There is a difference between formative assessments such as homework and quizzes that allow a teacher to plan instruction and spot-check student understanding versus a summative assessment to check student understanding on an entire instructional unit.²²⁷

The primary difference between classroom assessments and standardized tests is that standardized tests are summative assessments that focus on broad areas of learning, such as overall achievement in a subject matter rather than progress in a subject matter area.²²⁸ Standardized tests typically: 1) are summative; 2) cover varied content, over specific or general topics, across districts and states; 3) are created by a panel of experts; 4) will take several hours to complete; 5) are machine-scored; 6) are highly reliable; 7) are used to compare the normative score of the group with other test takers; and, 8) are used to determine general ability or achievement.²²⁹

There are several academic purposes that standardized tests claim to serve. First, single-subject survey tests might contain several subsets of tests used to check a student's achievement in a domain-specific content area, like reading.²³⁰ There are also standardized tests to assess current knowledge—standardized achievement tests—and readiness, which education administrators can use for classroom placement—standardized aptitude tests, career or educational interest inventories, and personality tests.²³¹ The academic purpose of a test depends on the specific standardized test administered to a student.

For example, education administrators use standardized achievement tests to identify the strengths and weaknesses of individual students and school districts.²³² Aptitude tests are used from early elementary school through

²²⁴ Kevin Seifert & Rosemary Sutton, *Ch. 16 Standardized and Other Formal Assessments, in INSTRUCTIONAL METHODS, STRATEGIES AND TECHNOLOGIES TO MEET THE NEEDS OF ALL LEARNERS* (2009), <https://granite.pressbooks.pub/teachingdiverselearners/chapter/standardized-and-other-formal-assessments-2/> [https://perma.cc/X37A-PZ5A].

²²⁵ DURWIN & REESE-WEBER, *supra* note 109, at 535.

²²⁶ *Assessment Glossary*, NAT'L COUNCIL ON MEASUREMENT IN EDUC., <https://www.ncme.org/resources/glossary> [https://perma.cc/RN75-CG3P].

²²⁷ DURWIN & REESE-WEBER, *supra* note 109, at 535.

²²⁸ *Id.*

²²⁹ *Id.*

²³⁰ *Id.* at 536, tbl.25.1.

²³¹ *Standardized Test*, GLOSSARY OF EDUC. REFORM (Nov. 12, 2015), <https://www.edglossary.org/standardized-test/> [https://perma.cc/3W43-SJKU].

²³² DURWIN & REESE-WEBER, *supra* note 109, at 536.

undergraduate and graduate admissions to make decisions about placement, or in the case of higher education, acceptance into a program of study.²³³ These tests also purport to assess future potential—the capacity to learn—in general or specific domains.²³⁴ Career and educational interest inventories are used more as guidance tools to assist high school and college students choose a career path or assist companies in determining who is a good candidate to hire.²³⁵ Finally, personality tests, which are limited in their educational use, as they are typically used to diagnose a clinical disorder, should only be used on individuals eighteen or older and can assess interests, attitudes, values, and patterns of behavior.²³⁶ In sum, the academic purpose of a standardized test can vary as widely as deciding to place a second-grader in advanced math to whether or not a certain career path may interest a senior in high school—depending on the standardized test being administered.

This Article presents an abolitionist view when it comes to standardized tests. An abolitionist bases their view on the argument that even with all of the collected data, standardized tests do not truly represent what a student knows. It is the abolitionist's view that standardized tests are extremely narrow in what they can measure. To use a hypothetical, say that an otherwise well-achieving student with high marks in all of her classes takes a standardized test the day after a family member passed away. The singular snapshot of that day, when the student is mourning the loss of someone close to her, could never come close to accurately measuring her knowledge.

Second, abolitionists believe the research-supported view that standardized tests are normed to white, middle-class culture.²³⁷ For instance, “students with LEP [limited English proficiency] may do poorly on a math achievement test not because they have poor math skills but because they had difficulty understanding the questions and answer options given in English.”²³⁸ Furthermore, at-risk students—students that the U.S. government defines as 1) living in a household where a mother has less than a high school education; 2) a household utilizing our social safety net for Supplemental Nutrition Assistance Program benefits; 3) a single-parent household; and/or 4) a home where the parents' primary language is not English—are all at risk of struggling in high stakes testing.²³⁹

There are alternatives to standardized tests for abolitionists, which may seem cumbersome at first. The alternative proposed by this author is that teachers grade students on their progress throughout the semester. Abolitionists

²³³ *Id.* at 537.

²³¹ *Id.*

²³⁵ *Id.*

²³⁶ *Id.*

²³⁷ Mary Catherine O'Connor, *Aspects of Differential Performance by Minorities on Standardized Tests: Linguistic & Sociocultural Factors*, in *TEST POLICY AND TEST PERFORMANCE: EDUCATION, LANGUAGE, AND CULTURE* 128 (Bernard R. Gifford ed., 1989).

²³⁸ DURWIN & REESE-WEBER, *supra* note 109, at 550.

²³⁹ *Id.*

do not believe that professional test writers exist.²⁴⁰ A teacher can observe their student for the entire year, through good and bad times.

Teachers should be able to collect a portfolio of their students' work and make joint decisions on placement while sitting down with each student, their parent(s) or guardian(s), and guidance counselor while reviewing the student's progress and what they showed they learned over the year. While there is an appeal to the pro-standardized test angle—that these tests provide something to differentiate individuals and do it *en masse*—abolitionists believe that not knowing what a student truly understands can be detrimental to their educational engagement.

While there are inherent flaws in the portfolio-based assessment, such as a teacher's racial, gender, or other biases, it is likely the participation of many stakeholders will balance out these flaws. Other flaws may include discrepancies in each teacher's ability to write effective summative and formative assessments or a student's disengagement in school for other reasons unrelated to the classroom teacher. While each of these flaws is most likely present in most educational settings, the author believes that, while there is no such thing as an ideal way of examining a student's assessment, the collection of longitudinal data and the inclusion of multiple stakeholders will mitigate these issues to the greatest extent possible.

IV. CONCLUSION

At-risk students need to be elevated, not just shuffled into some box with a number on it, and then tracked into a course in school where they cannot show their knowledge. To abolitionists, a cumulative assessment that combines elements of formative assessments and summative assessments along with accounting for the student's personality, willingness, eagerness to learn, and take on challenges, as well as the obstacles that they have overcome, is a much more holistic way to look at students. For too long standardized tests have perpetuated an undemocratic system of test-taking where affluent, white students dominate test scores. It is time for educators, students, and parents to reclaim education and actualize what Horace Mann said when he stated, "Education, then, beyond all other devices of human origin, is a great equalizer of the conditions of men—the balance wheel of the social machinery."²⁴¹

²⁴⁰ Peter Greene, *The Writer Who Couldn't Answer Standardized Test Questions About Her Own Work (Again)!*, FORBES (Apr. 28, 2019, 4:03 PM), <https://www.forbes.com/sites/petergreene/2019/04/28/the-writer-who-couldnt-answer-standardized-test-questions-about-her-own-work-again/?sh=778bf7634e8c> [<https://perma.cc/R8MV-DJ6C>].

²⁴¹ Roslin Growe & Paula S. Montgomery, *Educational Equity in America: Is Education the Great Equalizer?*, 25 PRO. EDUCATOR 23, 23–24 (2003), <https://files.eric.ed.gov/fulltext/EJ842412.pdf> [<https://perma.cc/G2T3-3J4H>].

**APPENDIX A: HISTOGRAMS BASED ON STATISTICS PROVIDED
BY THE COLLEGE BOARD FOR THE PSAT 8/9 (2018)**

Figure 1:²⁴²

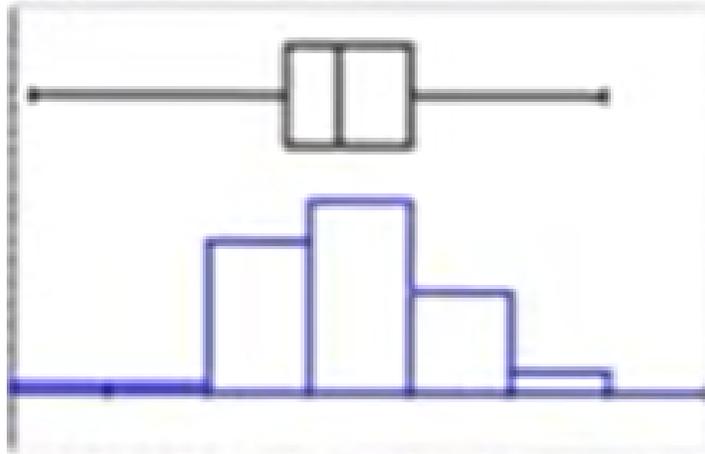
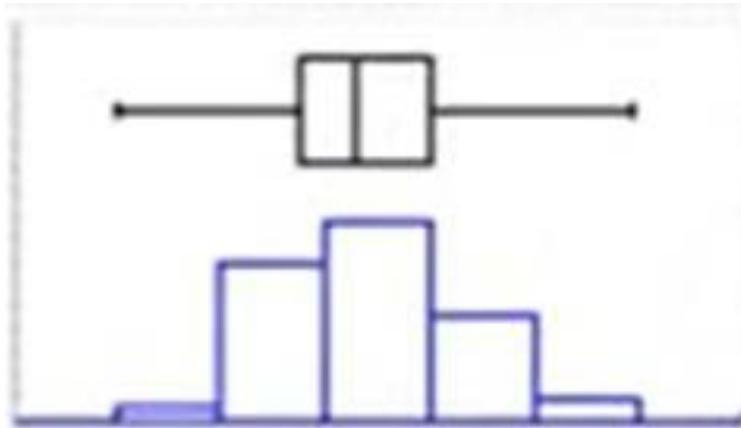


Figure 2:²⁴³



²⁴² COLL. BD., *supra* note 126 (This histogram portrays the extremes of the minimum and maximum of the data set. Here we assume that the minimum is 200 and the maximum is 1440, which means that we cannot see if the histogram is skewed to the right, left, or is normally distributed).

²⁴³ COLL. BD., *supra* note 126 (This histogram uses 240 as the minimum and 1440 as the maximum. As stated above in note 242, the data cannot be interpreted as skewed to the right, left, or is normally distributed).