

Teach to the Student, Not the Test

Bonnie K. Robinson

The University of North Carolina at Charlotte

Alicia Dervin

The University of North Carolina at Charlotte

National education reform has led to increased accountability and high-stakes testing in primary and secondary grades. Testing has become a billion-dollar industry in the United States, and many schools have created a culture of teaching to the test, while eliminating or reducing instruction of non-tested subjects. Standardized assessments are used to evaluate teachers and schools. However, validity can be questioned when significant gaps in student performance continue to exist between racial subgroups. The purpose of this article is to examine progress made in closing the Black-White achievement gap, by analyzing fourth and eighth graders' achievement on NAEP assessments from 2005 to 2017. The author suggests reducing the number of standardized assessments, and replacing test preparation curriculum with culturally relevant instruction.

Keywords: standardized testing, high stakes testing, culturally relevant pedagogy

Over the past fifty years, trends on national assessments have linked low academic achievement with children of color and low-income students (Hedges & Nowell, 1999; Koretz, 2017; Moore & Lewis, 2012). In an attempt to track and increase student achievement, a series of national education reforms have been created, placing a greater emphasis on high stakes testing in the United States (Koretz, 2017; Zhao, 2018). In 1965, President Johnson signed the Elementary and Secondary Education Act as part of his “War on Poverty”. In an effort to increase success for low-income students, the Johnson administration created Head Start preschools and Title I funding (Ellis, 2007). Twenty years later, the Reagan administration published *A Nation at Risk* in 1983, which provided a litany of alarming statistics on America’s failing education. The landmark report urged for reform, including more rigorous and measurable standards (Graham, 2013). No Child Left Behind (NCLB) was enacted nearly two decades later, federally requiring all states to administer standardized tests in reading and math in grades 3-8, and once in high school. States were required to report test data on a variety of subgroups based on race, income, and ability (Ellis, 2007).

In 2009, the Obama administration unveiled Common Core State Standards and Race to the Top grants, which encouraged states to adopt common standards and assessments. In 2015, the Every Student Succeeds Act (ESSA) was signed into law. Among other things, ESSA measures student progress of high standards through annual statewide assessments (Sanchez & Turner, 2017). While every federal reform has been created with an intention of creating equal opportunity and improving quality of education, the increased accountability has led to a testing culture in schools where teachers are pressured to teach to the test, rather than the individual students. This article seeks to answer the questions: (a) how has increased accountability through high stakes testing impacted the Black-White achievement gap, and (b) has the Black-White achievement gap narrowed since NCLB?

Literature Review

Standardized testing has become highly profitable, surging to a \$2 billion industry in the United States. The testing industry is monopolized by four major companies: Harcourt, McGraw-Hill, Houghton Mifflin, and Pearson (Kamenetz, 2015). In 2017, Pearson alone delivered 25.3 million standardized tests online, and 20.4 million paper-based assessments for K-12 students in the U.S. (Davis & Molnar, 2018). Education reforms have drastically increased the frequency of standardized testing. Since No Child Left Behind, every public-school student in grades 3-8 must take at least two standardized tests per year, however, the average student will take ten or more standardized tests each year (Lazarin, 2014). The Center for American Progress published a report of research conducted in 14 districts across seven states for the 2013-2014 school year (Lazarin, 2014). According to Lazarin (2014), “Students are tested as frequently as twice per month and an average of once per month. [...] students take as many as 20 standardized assessments per year and an average of 10 tests in grades 3-8” (p.3). Despite a high frequency of standardized testing, the amount of time spent testing in the districts that were studied was less than 2% of instructional time. However, researchers found a culture of testing that placed more emphasis on test preparation than actual learning (Lazarin, 2014).

Several states have implemented additional testing for designated grades. Studies from North Carolina Department of Public Instruction (Guindon, Huffman, Socol, & Takahashi-Rial, 2014) reveal that third graders spend a minimum of 13 hours each year on state and federal assessments. North Carolina mandates mClass assessments for all K-3 students, which are administered to students individually. This one-to-one testing means students will not receive instruction while the educator individually tests their classmates. According to Guindon et al. (2014), “One LEA [Local Education Agency] estimated that teachers lose 45 hours of instruction per year due to mClass” (p. 4). In addition to national and state assessments, students spend significant time completing district benchmarks and common formative assessments. Kamenetz (2015) found that New York schools designated seven weeks of their 36-week school year for testing. This included practice tests and mock assessments to increase students’ testing stamina. Teachers spend hours of instruction preparing students with test preparation material, including how to best answer multiple choice questions (Koretz, 2017; Zhao, 2018).

Not only are standardized assessments changing how teachers instruct students, they also impact what content is taught. Tested subjects such as math and reading receive greater emphasis in instructional time, while the arts, social studies, and sciences are neglected (Rose, 2014). Low-performing students are often encouraged to take test-related courses rather than electives (Greene, 2018). In an effort to increase test scores, children of color are more likely to receive drill-and-skill teaching through rote memorization (Kohn, 2000; Zhao, 2018). Milner (2013) cautions against scripted and narrowed curriculum reform, particularly in urban settings, as it limits authentic learning and makes it more difficult for teachers to respond to individual classroom needs. This has future implications, as evidenced by 2015-2016 graduation rates, where Black and Hispanic populations were 5-10 percentage points below the national graduation rate (“Data: U.S. Graduation”, 2017).

In global comparisons, administering annual standardized tests is unique to the United States, particularly at the elementary level. Many nations around the world administer national assessments periodically, as is the case in Australia, where students are tested in grades 3, 5, 7, and 9, or Canada, where students are assessed in grades 3, 6, 9, and 12 (Darling Hammond, 2017). Many high performing countries, including Finland, Japan, and China, do not administer any high stakes tests to elementary or lower secondary students (Darling Hammond, 2017; Rotberg, 2006). The Ministry of Education in Japan monitors the nation’s education performance by conducting assessments to a sample of sixth and ninth grade students (Tucker, 2011). In

Finland, teachers are highly trained to collect small data at the classroom level. Finnish students do not take lengthy standardized assessments, yet they consistently outperform U.S. students on international math and reading assessments (Sahlberg, 2018; Tucker, 2011).

Another commonality between these countries is that testing is not used as a form of teacher evaluation. Assessments are given to a sampling of students and test results are used as an educational quality control, rather than an accountability tool for schools or teachers (Darling-Hammond, 2017; Tucker, 2011). This is in stark contrast to practices in the United States, where the effectiveness of educators is often based on test scores. Koretz (2017) theorizes multiple reasons why evaluating teachers based on standardized assessments is a poor practice. Value added models are taken into account to determine growth of a student, though students predicted scores are based solely on a previous assessment score, rather than outside factors. Also, teacher test scores vary from year to year, which is more indicative of class dynamics rather than individual teacher effectiveness (Koretz, 2017). In the nations mentioned above, standardized assessments are used to measure student growth, and serve as gateways for students to access greater educational opportunities. This increases student and family responsibility, rather than placing accountability on the teacher alone (Darling-Hammond, 2017; Tucker, 2011).

The validity of standardized test scores must be questioned when academic achievement can be predicted by parent socioeconomic status. Longitudinal studies have found a strong correlation between test scores and income levels (Battler & Lewis, 2002; Hedges & Nowell, 1999; Zhao, 2018). Many states across the nation are evaluating schools based on test scores, though studies repeatedly correlate school letter grades with wealth (Ableidinger, 2015). Students' school experiences vary greatly based on geographic location and economic status. Students in urban schools are more likely to be educated by novice teachers using more scripted curriculum, and have access to fewer resources (Milner, 2013; Moore & Lewis, 2012). Considering these factors, it is not surprising that urban students were outperformed by suburban and rural students on 2017 National Assessment of Education Progress (NAEP) reading and math assessments (National Center for Education Statistics, 2018a; National Center for Education Statistics, 2018b). Validity can also be questioned when one examines the bias in standardized testing. Townsend (2002) writes, "Standardized tests have long been considered unfair and biased against students from ethnic minority and/or impoverished backgrounds because these tests are based in large measure on the experiences of middle-class European Americans" (p. 223).

These factors have led to perpetual test scoring discrepancies, most notably between Black and White students. It was the racial discrepancies in academic achievement that led to authorization of No Child Left Behind, which urged for an increase in accountability and tracking by subgroups. Rothert (n.d.) shares, "Statistics show that 12th-grade African American and Latino students have reading and math skills roughly equivalent to those of eighth-grade white students" (para. 3). Hedges and Nowell (1999) identify possible causes of the Black-White gap in test scores, including differences in social class, differences in family structure, and discrimination against Black students. They compared results from multiple standardized assessments of high school seniors, including NAEP trends from 1971 to 1996. According to their studies, Hedges and Nowell (1999) hypothesize, "The NAEP data suggest that, at the current rate of change, the gap in reading achievement will close in about 30 years and the gaps in mathematics and science achievement in about 75 years" (p. 130). While their study mentioned the impact of social class and family structure on Black families, they neglected issues of school reform that may increase equity in schools. The following section seeks to determine what progress has been made in the twenty years following Hedges and Nowell's research.

Methodology

In an effort to examine if racial disparities and achievement gaps have closed since increased testing accountability, one can examine scores between multiple subgroups on national assessments. NAEP tests are administered to representative samples of students across the United States in selected grades. Student performance is reported through scale scores and achievement levels. NAEP sets three achievement levels: Basic, Proficient, and Advanced. All data used in this article was collected from the online NAEP Data Explorer tool. Historically, academic achievement between White and Black students has seen the greatest disparity, which is why those races were selected for this study. Data was compared between fourth graders and eighth graders for the following years: 2005, 2009, 2013, 2017. The subjects (reading and math) and grades (fourth and eighth) were selected because those grades and subjects have been heavily tested since the implementation of No Child Left Behind. The range of years selected correlate with curriculum and assessment implementations that resulted from No Child Left Behind, Common Core Standards, and Every Student Succeeds Act.

Results

There are significant gaps between Black and White students scoring at or above proficient in math (see Figure 1 in Appendix A). White students consistently outperform Black students by 30 or more percentage points in each measure. The disparity between Black and White fourth graders in math is as follows: 34 percentage points in 2005, 35 percentage points in 2009, 36 percentage points in 2013, and 32 percentage points in 2017. Differences among Black and White eighth graders are 30 percentage points in 2005, 32 percentage points in 2009, 31 percentage points in 2013, and 31 percentage points in 2017.

Another way to view the data is by looking at cohorts of students, comparing fourth graders with eighth graders' scores four years later. When comparing the cohorts, the Black-White achievement gap narrowed by as much as five percentage points. However, it is also important to note that the overall proficiency between fourth and eighth graders dropped every year with both races, as much as 10 percentage points from the 2013 fourth graders to 2017 eighth graders. Among the eight samples of math tests, the highest percentage of Black students scoring at or above proficient was a mere 19%, compared to the highest percentage of White students achieving 54% proficiency.

Gaps in reading proficiency between Black and White students are only slightly smaller than math (see Figure 2 in Appendix B). The difference between Black and White fourth graders in reading is as follows: 28 percentage points in 2005, 26 percentage points in 2009, 28 percentage points in 2013, and 27 percentage points in 2017. Differences among Black and White eighth graders are 27 percentage points in 2005, 27 percentage points in 2009, 29 percentage points in 2013, and 27 percentage points in 2017. When examining the cohorts of students, percentage points fluctuated minimally, with increases by one point and a decrease by three percentage points. Similar to math assessment results, the highest number of Black students scoring at or above proficient was a mere 20%, compared to a maximum percentage of 47% for White students. In other words, only one in five Black students are reading at what the nation considers a proficient level.

Discussion

The purpose of this study was to examine the impact of high stakes testing on narrowing the Black-White achievement gap. NAEP data was analyzed from fourth graders and eighth graders in math and reading, during the years 2005, 2009, 2013, and 2017. As evidenced by the figures above, there has been a steady increase in proficiency levels in both math and reading

since 2005. However, the Black-White achievement gap persists, and there has been no significant decrease in these gaps since No Child Left Behind was signed into law in 2002. Increased testing has led to an increase in test prep materials and teaching to the test, and four out of five Black students are scoring below proficient in math and reading assessments. Despite efforts to close the achievement gap, the age of increased accountability and high stakes testing has made no significant progress towards closing the gap. Between the years 2005 and 2017, there has been a two-point decrease in achievement gaps between fourth grade math, a one-point increase in gaps between eighth grade math, a one-point decrease in gaps between fourth grade reading, and no difference in eighth grade reading.

Ladson-Billings (2006) argued that, rather than an achievement gap, a more accurate term would be educational debt. The disparity among standardized test scores is far more complex than education reform. One must factor in the historical debt, which is directly tied to economic and sociopolitical debt that people of color, particularly African Americans, have disproportionately been given from centuries of slavery, legal segregation, and discrimination (Ladson-Billings, 2006). Considering these factors, it is a drastic over-simplification to believe that an increase in standardized tests and accountability would provide equity. On the contrary, it is children of color who most often suffer from ramifications of low-test scores.

Recommendations

In order to provide equity and increase student achievement for all races, the best solution is for educators to adopt culturally relevant teaching. Gloria Ladson-Billings (1995) developed the conceptual framework of culturally relevant pedagogy, establishing the following criteria for educators: develop students academically, nurture cultural competence, and develop sociopolitical consciousness. This approach gives teachers the flexibility and autonomy to examine the needs of their unique classroom, while engaging students in authentic learning. Throughout the pendulum swings of various education reforms in the past decades, it is clear that there is no one-size fits all program or curriculum (Zhao, 2017). The previous sections of this article reveal that narrowing curriculum, test preparation, and increased testing have not closed achievement gaps between Black and White students. Culturally relevant pedagogy has increased learning outcomes, particularly for Black students, beyond test scores (Moore & Lewis, 2012).

Before educators can implement culturally relevant pedagogy, they must first examine their own cultural bias and views of society (Schmeichel, 2012). As of 2017, 80% of public school teachers in the United States were White females. This teaching population is not representative of the majority of students in U.S. public schools, who are now children of color (Strauss, 2014). This cultural divide makes it all the more necessary to address cultural bias among staff and preservice educators. An emphasis on diversity and equity training should begin at the university level with systemic implementation of culturally relevant pedagogy. Preservice teachers often lack experience interacting with diverse student populations. Jackson and Boutte (2018) write, "Close examination of many teacher education programs reveals that the focus on issues of equity and CRP is typically superficial and not supported by practices, instruction, curriculum, policies, and dispositions of teacher educators" (p. 87). Muschell and Roberts (2011) provide an overview of how preservice teachers receive culturally relevant training through a series of course assignments at Georgia College and State University. Students begin by examining their cultural awareness through creations of cultural collages and autobiographies, then move into exploring social justice topics through children's literature and multicultural children's literature. It is also critical to carefully select field experiences that focus on diversity (Ellerbrock, Cruz, Vásquez, & Howes, 2016).

In addition to preservice teacher education at the university level, districts should provide culturally relevant training to all staff. Funds could be reallocated from the billion-dollar testing industry to provide ongoing professional development. Universities can offer summer programs for teachers, or provide extended trainings to give teachers time to collaborate, engage in reflective discourse, and write curriculum and units of study that best meet the needs of their students. Staff members can work together in conjunction with students, families, and community members to support student achievement. One way to ensure teachers receive ongoing training in culturally relevant pedagogy is to require it as a part of licensure and teacher renewal. States all across the country determine licensure requirements, mandating teachers spend a certain number of hours in professional development courses over the duration of several years. Continuing education credits in culturally relevant pedagogy should be as important as courses in math, reading, technology, etc.

Accountability is still necessary to determine if equity in education is achieved, however, progress can be measured in alternative ways beyond traditional standardized assessments. One possibility is replacing testing with portfolio-based assessments, where students demonstrate learning through a variety of tasks. Student portfolios are evaluated at 28 New York secondary schools, and these schools have higher graduation rates and better college-retention rates than surrounding schools with similar demographics (Kamenetz, 2015). Another creative alternative to standardized assessments is conducting schoolwide inspections. In place of national assessments, Scotland has a system of inspections where government inspectors examine student work and interview students and staff (Kamenetz, 2015). As mentioned in the literature review section of this paper, the United States is one of few countries that has standardized assessments for all students, starting in elementary school. The United States could adopt practices from other high performing countries and reduce or eliminate standardized tests. Sampling of students could continue to be assessed on national and international assessments to determine disparities in scores among subgroups of race and ability.

Conclusion

This article addressed the questions: (a) how has increased accountability through high stakes testing impacted the Black-White achievement gap, and (b) has the Black-White achievement gap narrowed since NCLB? Increased standardized assessments have not generated desired results. Significant gaps of 25 or more percentage points remain between Black and White students scoring at or above proficient on reading and math NAEP assessments. In place of standardized assessments, we suggest reducing the number of high stakes testing and focusing on culturally relevant pedagogy to best deliver instruction. Accountability may occur through portfolio-based assessments, school inspections, or sampling for national assessments. The focus of learning should be on the individual child, rather than on the test.

References

- Ableidinger, J. (2015). A is for affluent. *Public School Forum of North Carolina*. Retrieved from <https://www.ncforum.org/wp-content/uploads/2016/10/A-is-for-Affluent-Issue-Brief-Format.pdf>
- Battle, J., & Lewis, M. (2002). The increasing significance of class: The relative effects of race and socioeconomic status on academic achievement. *Journal of Poverty*, 6(2), 21–35. doi: 10.1300/J134v06n02_02
- Darling-Hammond, L. (2017). *Empowered educators: How high performing systems shape teaching quality around the world*. San Francisco, CA: Jossey Bass.
- Davis, M. R. & Molnar, M. (2018). Educators carefully watch Pearson as it moves to see K-12 curriculum business. *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2018/03/07/educators-carefully-watch-pearson-as-it-moves.html>
- Data: U.S. graduation rates by state and student demographics. (2017). *Education Week*, 37(15). Retrieved from <https://www.edweek.org/ew/section/multimedia/data-us-graduation-rates-by-state-and.html>
- Ellerbrock, C., Cruz, B., Vásquez, A., & Howes, E. (2016). Preparing culturally responsive teachers: Effective practices in teacher education. *Action in Teacher Education*, 38(3), 226–239. doi: 10.1080/01626620.2016.1194780
- Ellis, C. R. (2007). No child left behind – A critical analysis: "A nation at greater risk". *Curriculum and Teaching Dialogue*, 9(1), 221-233, 334. Retrieved from <http://search.proquest.com/docview/230423291/>
- Graham, E. (2013). 'A Nation at Risk' turns 30: Where did it take us? *NEA Today*. Retrieved from <http://neatoday.org/2013/04/25/a-nation-at-risk-turns-30-where-did-it-take-us-2/>
- Guindon, M., Huffman, H., Socol, A.R., & Takahashi-Rial, S. (2014). *How much testing is taking place in North Carolina schools at grades K-12? An analysis of federal, state, and local required assessments*. North Carolina Department of Public Instruction. Retrieved from <http://www.dpi.state.nc.us/docs/intern-research/reports/testing2014.pdf>
- Hedges, L., & Nowell, A. (1999). Changes in the black-white gap in achievement test scores. *Sociology of Education*, 72(2), 111–135. doi: 10.2307/2673179
- Jackson, T., & Boutte, G. (2018). Exploring culturally relevant/responsive pedagogy as praxis in teacher education. *The New Educator*, 14(2), 87–90. doi: 10.1080/1547688X.2018.1426320
- Kamenetz, A. (2015). *The test: why our schools are obsessed with standardized testing but you don't have to be*. New York, NY: PublicAffairs.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Koretz, D. (2017). *The test charade: Pretending to make schools better*. Chicago, IL: The University of Chicago Press.
- Ladson-Billings, G. (1995) Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491.

- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3–12.
- Lazarin, M. (2014). *Testing overload in America's schools*. Center for American Progress. Retrieved from <https://cdn.americanprogress.org/wp-content/uploads/2014/10/Lazarin-OvertestingReport.pdf>
- Milner, R. H. (2013). Scripted and narrowed curriculum reform in urban schools. *Urban Education*, 48(2), 163–170. doi: 10.1177/0042085913478022
- Moore, J. L. & Lewis, C. W. (Eds.) (2012). *African American students in urban schools: Critical issues and solutions for achievement*. New York, NY: Peter Lang Publishing
- Muschell, L., & Roberts, H. (2011). Bridging the cultural gap: One teacher education program's response to preparing culturally responsive teachers. *Childhood Education*, 87(5), 337–340. doi: 10.1080/00094056.2011.10523209
- National Center for Education Statistics. (2018a). *National Assessment of Educational Progress (NAEP) Reading Report Card*. Retrieved from https://www.nationsreportcard.gov/reading_2017/nation/achievement
- National Center for Education Statistics. (2018b). *National Assessment of Educational Progress (NAEP) Math Report Card*. Retrieved from https://www.nationsreportcard.gov/math_2017/nation/achievement?grade=4
- Rose, M. (2014). School reforms fails the test. Retrieved from <https://theamericanscholar.org/school-reform-fails-the-test/#.W8yoWNMrLOQ>
- Rotberg, I. (2006). Assessment around the world. *Educational Leadership*, 64(3), 58–63.
- Rothert, C. (n.d.). Achievement gaps and No Child Left Behind. *National Center for Youth Law*. Retrieved from <https://youthlaw.org/publication/achievement-gaps-and-no-child-left-behind/>
- Sahlberg, P. (2018). *FinnishED leadership: Four big, inexpensive ideas to transform education*. Thousand Oaks, CA: Corwin.
- Sanchez, C., & Turner, C. (2017). Obama's impact on America's schools. Retrieved from <https://www.npr.org/sections/ed/2017/01/13/500421608/obamas-impact-on-americas-schools>
- Schmeichel, M. (2012). Good teaching? An examination of culturally relevant pedagogy as an equity practice. *Journal of Curriculum Studies*, 44(2), 211–231. doi: 10.1080/00220272.2011.591434
- Strauss, V. (2014). For first time, minority students expected to be majority in U.S. public schools this fall. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2014/08/21/for-first-time-minority-students-expected-to-be-majority-in-u-s-public-schools-this-fall/?utm_term=.1e90a1ab396e
- Townsend, B. L. (2002). Testing while black. *Remedial and Special Education*, 23(4), 222. Retrieved from <https://librarylink.uncc.edu/login?url=https://search-proquest-com.librarylink.uncc.edu/docview/236324173?accountid=14605>
- Tucker, M. S. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Cambridge, MA: Harvard Education Press.
- Wilson, W. (1978). The declining significance of race. *Society*, 15(2), 56–62.

doi: 10.1007/BF03181003

Zhao, Y. (2018). *What works may hurt: Side effects in education*. New York, NY: Teachers College Press.

Appendix A

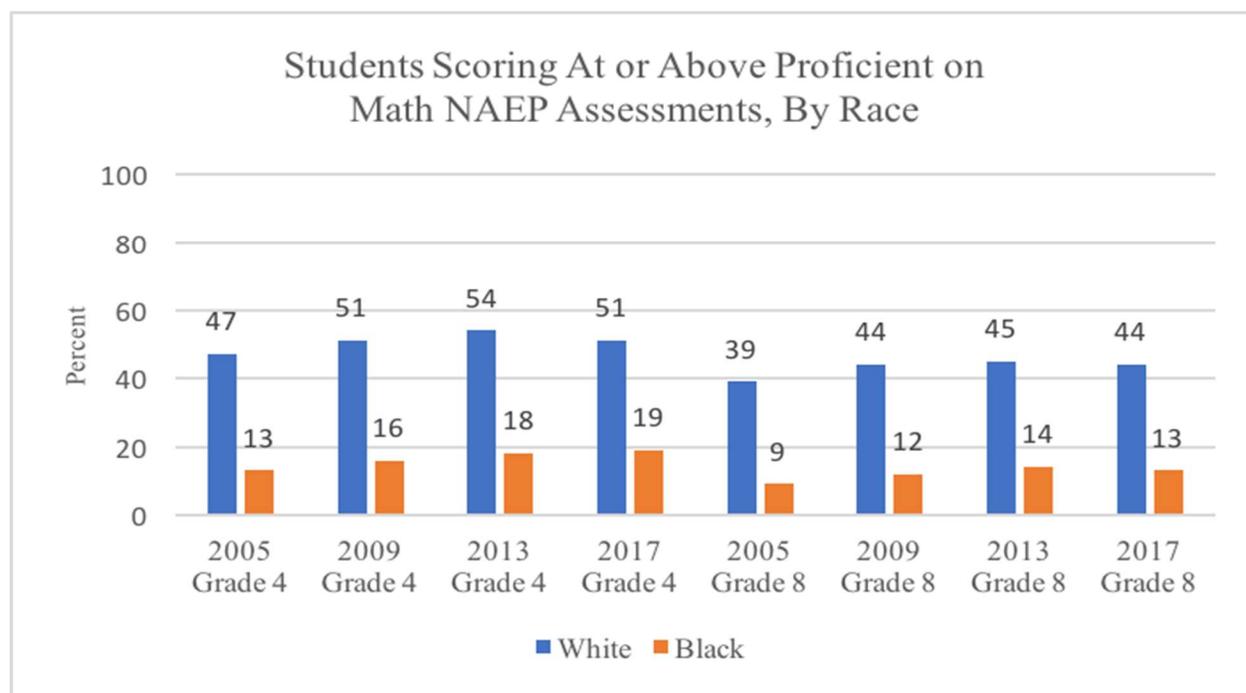


Figure 1. Students scoring at or above proficient in math NAEP assessments. Adapted from U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2009, 2013, and 2017 Math Assessments.

Appendix B

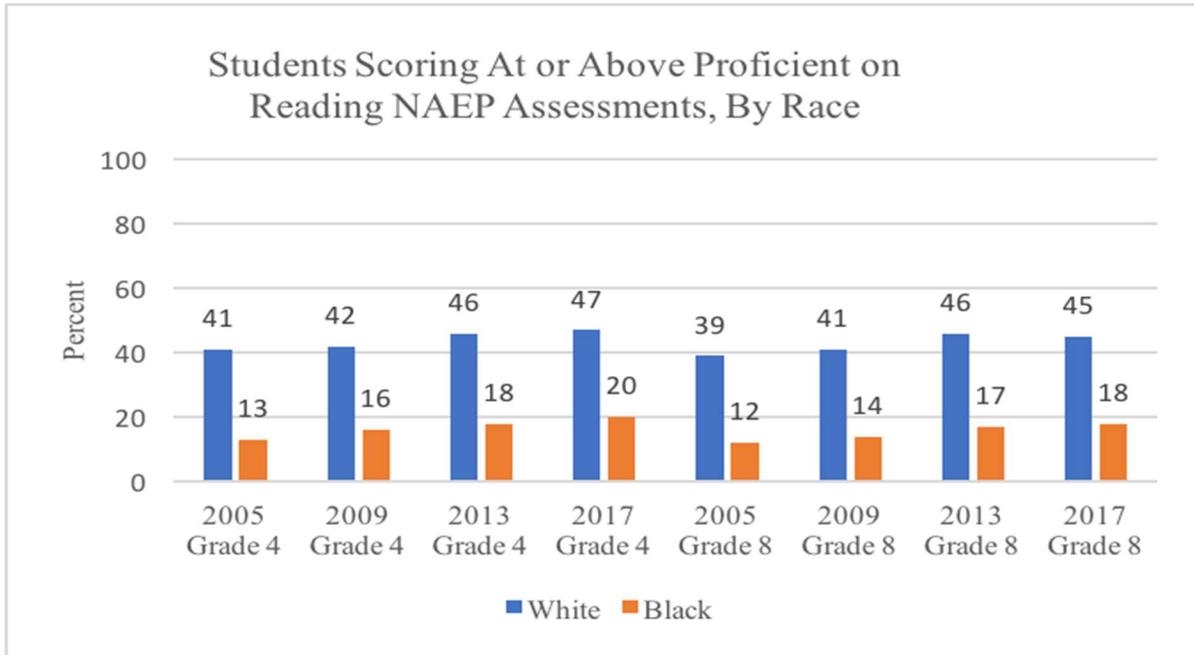


Figure 2. Students scoring at or above proficient in reading NAEP assessments. Adapted from U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005, 2009, 2013, and 2017 Reading Assessments.