

The Effectiveness of Title I: Synthesis of National-level Evidence from 1966 to 2013

Sonia Sousa*, David Armor

School of Public Policy, George Mason University, USA

Copyright © 2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This paper reviews prior research to assess the effectiveness of Title I in closing the achievement gaps of disadvantaged students vis-à-vis their non-disadvantaged counterparts. A research synthesis approach is adopted to summarize national assessments of Title I conducted between 1966 and 2011. These analyses are supplemented by the authors' analysis of NAEP data from 1990 to 2013. There is no evidence that early Title I programs significantly reduced achievement gaps nationwide. Studies following NCLB implementation show modest closure of grade 4 gaps of about 0.2 of a standard deviation. Given the modest academic gains attributable to Title I, and considering that the program costs about \$15 billion per year, the authors conclude that Title I programs have not been cost effective in closing the achievement gaps.

Keywords Title I, NCLB, Education, Achievement Gaps, Research Synthesis

1. Introduction

Established as part of the Elementary and Secondary Education Act of 1965 (ESEA), Title I is a U.S. compensatory education program that provides federal financial assistance to elementary and secondary schools, mostly public schools, with a high proportion of children from low-income families. Since its inception in 1965, Title I has been the largest single program in the U.S. Department of Education, accounting for close to 40 percent of the Department of Education's total K-12 budget in recent years. In 2012, its annual funding was about \$14.5 billion, and it reached over 23 million school children.

Title I was established with the original goal of improving the educational attainment of children in poverty. The No Child Left Behind Act of 2001 (NCLB) made this goal more concrete by establishing a requirement to attain 100 percent proficiency for all students by the 2013-14 school year. One implication of the NCLB 100-percent-proficiency target is that Title I aims to raise the academic achievement of

disadvantaged students to match non-disadvantaged students, thereby closing achievement gaps by the 2013-14 school year.

This paper seeks to synthesize the national-level evidence on the effectiveness of the overall Title I compensatory program in closing the achievement gap between poor and non-poor students nationwide. The paper also discusses the costs of that contribution.

1.1. Title I Characteristics

One key characteristic of Title I is that funds are allocated to eligible schools based on the census estimates of children's poverty levels in the school district and the cost of education in the state. The poverty threshold and the scope of Title I programs have evolved over time. Currently, Title I schools with at least 40 percent of children from low-income families are eligible to use Title I funds, along with other Federal, State and local funds, to put in place school-wide assistance programs designed to improve academic achievement of all students. Title I schools below the 40-percent threshold and those that choose not to operate a school-wide program can use Title I to fund targeted assistance programs for students who are failing, or at risk of failing, to meet the state's academic achievement standards. Targeted assistance programs should be designed to meet the needs of those students and should be developed in consultation with parents, school staff, and district staff. Title I program provides some guidelines but school districts and schools have great flexibility to decide where and how to focus the funds.

NCLB introduced some significant changes to the Title I characteristics which went into effect in the 2002-03 school year. Four of these changes are worth mentioning here given their implications for the program's goals, costs and participation rates: (a) 100-percent state proficiency as the academic achievement target, (b) school-specific interventions, (c) teacher quality, and (d) parental choice. First, NCLB established a clear achievement target: 100 percent state proficiency in core subjects for all students by the 2013-14 school year. As noted above, this means that, in

practice, Title I’s goal now goes beyond raising academic achievement of poor students – it aims to close the achievement gaps between poor and non-poor students by 2013-14. Second, schools and districts failing to make adequate yearly progress towards the state proficiency targets are identified as needing improvement and are subject to specific interventions designed to improve their performance and provide additional options to students. Third, NCLB requires that all teachers (including Title I teachers) of core academic subjects are highly qualified. Fourth, parents of students attending continuously failing schools must be given the option of obtaining supplemental educational services from an approved public or private provider chosen by the parents and funded by Title I.

1.2. Costs and Participation

Figure 1 shows the trends in Title I appropriations (adjusted for inflation) as well as the number of students served by Title I compensatory programs. The funds shown cover only the Title I compensatory program and do not include funds for implementing components and programs of NCLB other than those funded by Title I and related programs.

Between 1966 and 1990 funding fluctuated between \$6 and \$8.3 billion (in 2012 dollars), swinging to the lower end during the 1980s. Funding increased in the early 1990s reaching \$10 billion by 1992 and steadied around this figure until 2000. Funding grew again in the early 2000s exceeding \$14.5 billion by 2008. Allocations in connection with the American Recovery and Reinvestment Act (ARRA) pushed spending to \$15.5 billion in 2009 and \$15.3 in 2010,

returning to \$14.5 billion in 2012.

Participation rates were not recorded during the early years of Title I. Until 1980 there were only broad guidelines on how the funds were to be spent, and individual student participants were not identified. Starting with the 1981 reauthorization act, Title I recipients (either individuals or whole schools) were counted as participants. Many students were identified for special pull-out sessions that offered more intense instruction or tutoring than was available in regular classrooms. Student participation rates remained fairly stable at about 5 million students from 1980 to 1995. It rose steadily in the late 1990s and 2000s, reaching approximately 20 million students by 2005 and 24 million in 2012. The main reason for the increase in the rate of participation since the late 1990s was the reduction of the threshold for school-wide programs. Before 1995, if a school had 75 percent of its students below the poverty line, the entire school enrollment would be counted as program participants. That threshold was reduced to 60 percent in 1995-96, then to 50 percent, and it is currently at 40 percent. Naturally, reducing the threshold has led to very large increases in Title I enrollment, which currently stands at nearly one-half of the national K-12 student population.

Per capita expenditures have been fairly low, particularly after 1997. Prior to 1997, per capita funding (in 2012 dollars) was about \$1500 per student, but starting in 1997 per capita funding dropped to \$900, and it continued to drop to just over \$600 in 2012. This drop is due to lowering the threshold for school-wide programs as the change counts more students as participants.

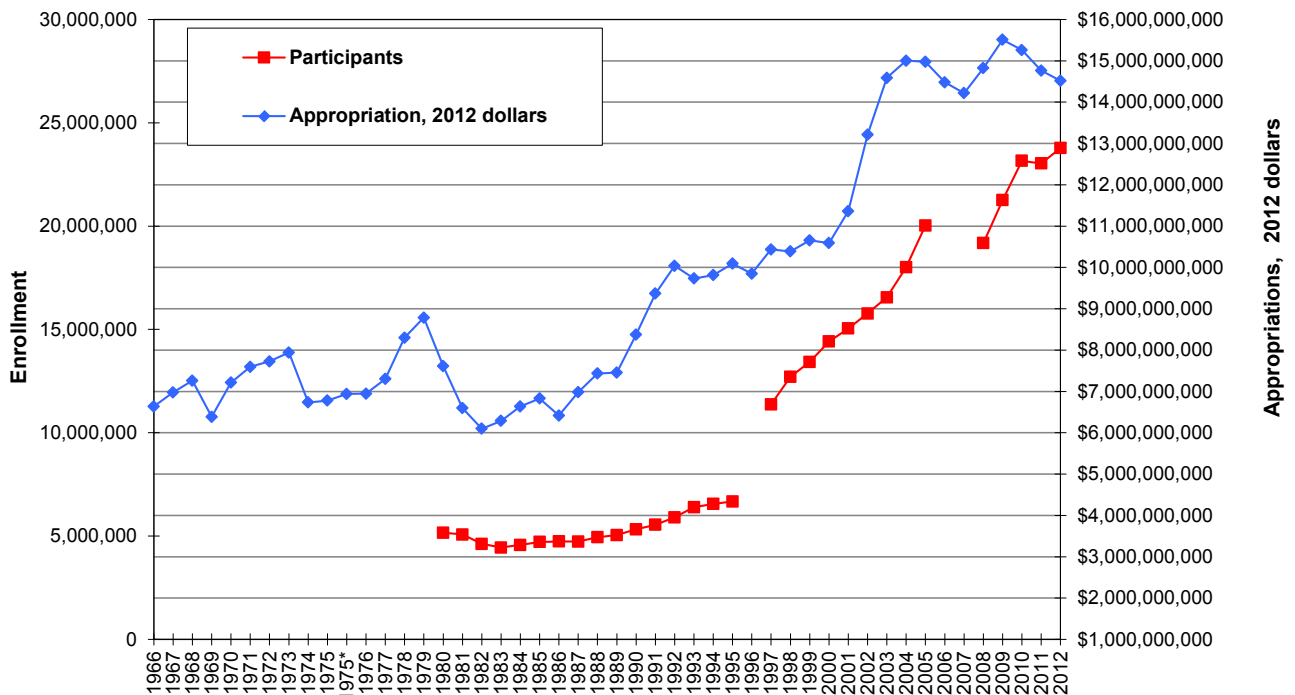


Figure 1. Trends in Title I Appropriations and Participation [1,2]

2. Materials and Methods

We conduct a research synthesis supplemented by an analysis of more recent longitudinal national-level NAEP data (1990–2013) to summarize the evidence on the outcome measure of interest to this paper: national-level achievement gap between poor and non-poor students. Research synthesis is a process through which we integrate the quantitative evidence provided by two or more research studies concerning a particular question, but not necessarily using an identical outcome measure across studies. A research synthesis lies between a literature review, which describes the authors' findings without presenting the quantitative results in a systematic manner, and a meta-analysis, which employs statistical methods to synthesize quantitative evidence in the form of an overall estimate of the effectiveness of a program/intervention. A research synthesis is particularly well suited to synthesize quantitative evidence based on a small number of studies and/or studies that use incomparable outcome measures [3]. We adopt a research synthesis approach, as opposed to a meta-analysis, because the disparate methodologies used to assess the effectiveness of Title I nationwide over the years makes it very difficult to produce an overall quantitative estimate of the effects of Title I on the achievement of below-poverty line students.

A rigorous assessment of Title I effectiveness would have to meet What Works Clearinghouse standards as set by the Institute for Education Sciences of the U.S. Department of Education. This would require selecting only studies employing randomized controlled trials (RCT) or true experimental designs. We did not apply this criterion because such rigorous evaluation studies have never been conducted for the overall Title I compensatory program nationwide, despite the fact that the Congress has specifically mandated three National Assessments of Title I since the late 1980s (more on this below). The most likely explanation for the absence of randomized designs is the near universal implementation of Title I, which makes it virtually impossible to find a randomly assigned control group that does not receive Title I services. A RCT study would require to randomly assign classrooms, schools, or school districts to Title I or non-Title I conditions. Given that Title I is a long standing nearly universal program, most would find ethical, political or legal objections to depriving eligible poor children from receiving Title I services [4,5].

2.1. Search Procedure and Criteria

We identified prior evaluations of Title I effectiveness using Google, Google scholar and ProQuest search engines, covering the period from 1970 to 2015. Depending on the engine search capability we employed the following keywords alone or in combination: Title I, no child left behind, impact, academic achievement. We scanned the title and abstract/executive summary to select only studies within

the scope of this research synthesis, which is defined by the following four criteria: (a) evaluations of Title I/NCLB as a whole, as opposed to evaluations of specific components; (b) evaluations of Title I/NCLB that are nationally representative, as opposed to assessments of state- or local-level effects of Title I; (c) evaluations that assess the effect of Title I/NCLB on students' academic achievement, as opposed to, for example, effects of Title I/NCLB on school spending, and (d) evaluations that look at students from disadvantaged poverty backgrounds as compared to their non-disadvantaged poverty counterparts. We then scanned the reference sections of the studies selected to explore other potential studies that would fit within the scope of this synthesis.

Studies assessing specific components of Title I, such as the impacts of remedial reading programs in [6] do not meet the requirements of this synthesis. Studies that assess the impact of specific NCLB provisions, as for example, [7-10] that assess school choice, supplemental educational services options, teacher quality and other accountability provisions are also outside the scope of this study. State- or local-level evaluations of Title I, such as [11-13] are also out of the scope of this synthesis. The study by Cascio and colleagues[14] is outside the scope because it addresses only the South and the outcomes studied are school spending and dropout rates. Studies that look at comparisons between groups defined by features other than poverty backgrounds are also outside of the scope of this synthesis. This is the case, for example of [15], which compares public vs non-public schools and high-standards vs non-high standards states.

2.2. Data

Our search yielded five peer reviewed studies and reports by the U.S. Department of Education that meet our data selection criteria stated above. These studies are: (a) the Borman and D'Agostino meta-analysis [16], which synthesizes evaluation studies conducted from 1966 to 1993; (b) the Prospects Study [17], which uses a large national sample of students and covers the 1991 – 1994 period; (c) the 1999 National Assessment of Title I and its follow-up studies [18-19], using national data from late 1980s to late 1990s; (d) the 2007 National Assessment of Title I and its follow-up studies [6,7,20,21], which cover the period from 1992 to 2007, and (e) the recent study by Dee and Jacob [22], which also analyses the period from 1992 to 2007. We supplement these sources with our own trend analyses of longitudinal national-level NAEP data from 2007 to 2013 to bring the academic achievement data up to date. Taken together, this research synthesis covers the period from 1966 to 2013.

3. Results

3.1. The Borman and D'Agostino Meta-analysis

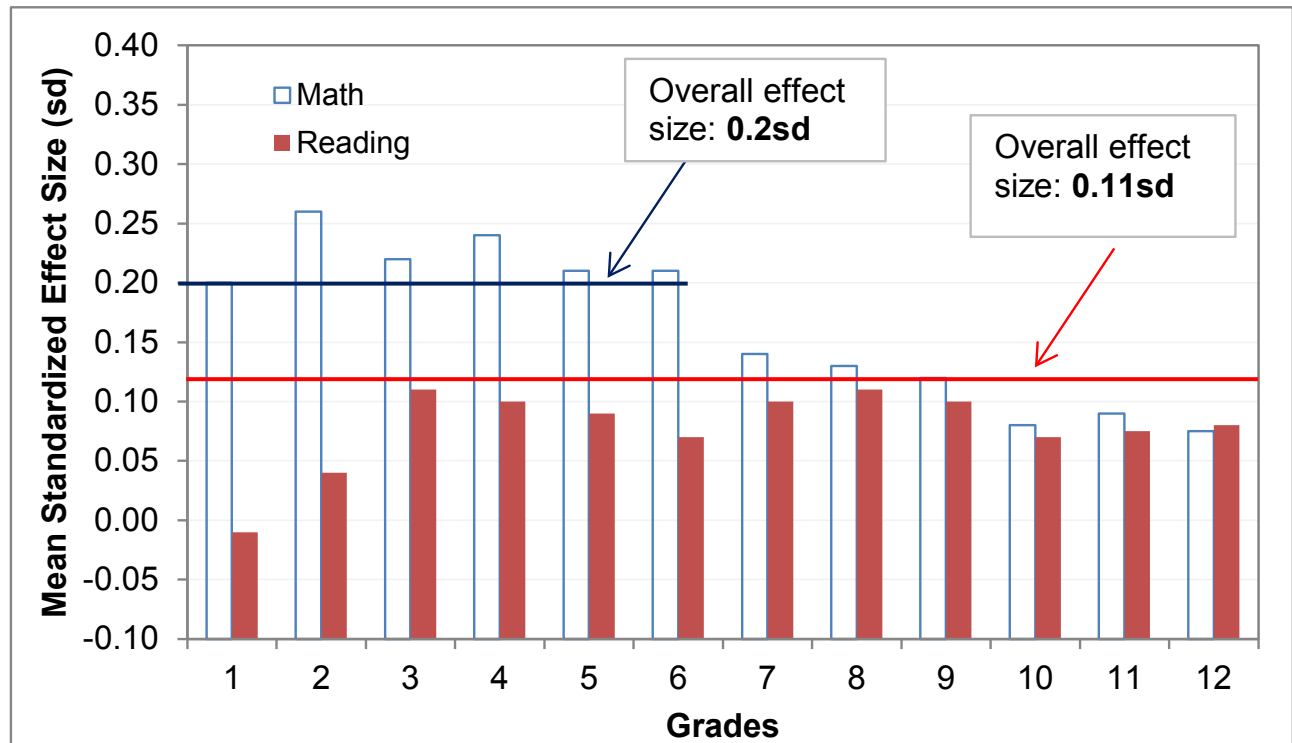


Figure 2. Mean Standardized Effect Sizes of Title I Participation from the Borman and D’Agostino Meta-Analysis (period covered: 1966 – 1993)

The Borman and D’Agostino meta-analysis [16] synthesizes 17 federal Title I evaluations conducted from 1966 to 1993. The outcome measure is the achievement gap between Title I participants and a control group. Figure 2 presents the mean standardized effect size, broken down by subject (math and reading) and by grade. Only studies employing a two-wave, pretest/posttest design [23] are included in the meta-analysis and reflected in the effect size of Figure 2. The standardized effect sizes represent the achievement gaps between Title I participants and non-participants as a fraction of one standard deviation (*sd*).

The results of this meta-analysis suggest three major findings. First, the overall effect of Title I over the 1966 to 1993 period is positive but modest in magnitude at only 0.11 *sd*. Second, for grades 1 to 6, Title I seems to have a stronger effect in math than in reading over this period. The effect sizes of Title I on students’ achievement in math range from 0.21 to 0.26 *sd* in grades 1 to 6, whereas the effect sizes for reading range from -0.01 to 0.11 *sd*. Third, the effect sizes of Title I in math decline significantly after grade 6 and resembles the effects in reading. Specifically, from grade 7 through grade 12, the effects of Title I in math decline to values ranging from 0.08 to 0.14 *sd*, while the effects in reading remain relatively similar to those in elementary grades, ranging from 0.08 to 0.11 *sd*. In short, the Borman and D’Agostino meta-analysis suggests that the effect of Title I is stronger in math programs and during the student’s elementary grades, although even here the effect is modest at about 0.2 *sd*.

3.2. The Prospects Study

The Prospects study [17] relies on a nationally representative sample of about 40,000 students from 364 schools, from grades 1, 3 and 7, from 1991 to 1994. Using the Comprehensive Test of Basic Skills (CTBS) to assess students’ academic achievement, the study estimates the effect of Title I by means of multivariate statistical techniques (namely, hierarchical linear models), to control for differences between Title I participants and nonparticipants on a set of student, family, and school characteristics.

Table 1 summarizes the effects broken down by number of years exposed to Title I, cohort, and subject area – math and reading. For each cohort and duration of exposure to Title I, the Prospects study estimates two related outcome measures: (a) achievement gaps, or the achievement score differences between Title I participants and nonparticipants and (b) changes in the achievement gap between participants and nonparticipants over the 1991 – 1994 period. We report the significant results of both outcome measures as standardized effect sizes. We obtained the standardized effects by dividing the scale-score points by the standard deviation of 45 scale-score points, the mid-point of the standard deviation range reported in the study. To illustrate for the 1st grade cohort in math, -0.64 *sd* means that students in 1st grade exposed to Title I for one year score, on average, approximately two-thirds of a standard deviation in math below those who are not exposed to Title I. The entry for over-time gap change means that the 1st grade math score gap of -0.64 *sd* does not significantly change over the 1991 – 1994 period.

Table 1. Summary of Effects of Title I Participation Detected by the Prospects Study (Period Covered: 1991 – 1994)

Academic Subject and Years of exposure to Title I	1 st Grade		3 rd Grade	
	Achievement gaps	Over-time gap change	Achievement gaps	Over-time gap change
Math (CTBS standard test)				
1 year	-0.64	not significant	-0.49	not significant
2 years	-0.62	not significant	-0.82	not significant
3 years	-0.60	not significant	-0.64	not significant
4 years	not applicable	not applicable	-1.02	0.11
Reading (CTBS standard test)				
1 year	-0.62	not significant	-0.38	not significant
2 years	-1.13	not significant	-0.53	-0.10
3 years	-1.38	not significant	-0.44	-0.08
4 years	not applicable	not applicable	-0.76	not significant

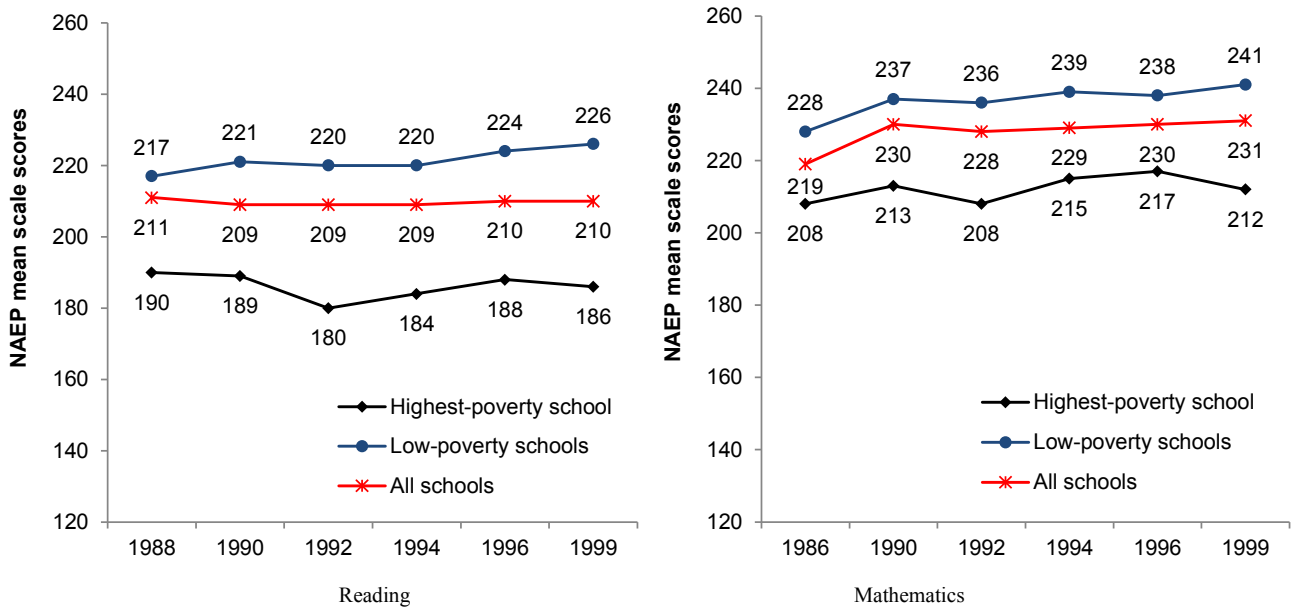
The results of the nationally representative Prospects study suggest three major findings. First, students exposed to Title I between 1991 and 1994 score, on average, considerably lower than students who did not receive Title I assistance. In both academic subjects and cohorts, participant students show moderate to large¹ negative achievement gaps (vis-à-vis nonparticipant students), on average, after being exposed to Title I for a minimum of one year and up to four years. For example, after three years of Title I exposure, the 1st-grade cohort scores, on average, 0.6 *sd* lower in math and 1.4 *sd* lower in reading compared to scores of students never exposed to Title I. For the 3rd-grade cohort, after 3 years of Title I exposure, on average, participant students score below nonparticipant students by 0.64 *sd* in math and by 0.44 *sd* in reading. Second, the negative achievement gaps are wider among students who have more years of Title I exposure than among students who have less years of Title I exposure. Third-graders with four years of Title I assistance show very large achievement gaps (vis-à-vis nonparticipants) in both math (-1 *sd*) and reading (-0.8 *sd*); whereas the negative achievement gaps after one year of Title I exposure are moderate to large in both math (-0.5 *sd*) and reading (-0.4 *sd*). Third, the negative achievement gap between participants and nonparticipants remain relatively unchanged over time between 1991 and 1994. The over-time gap changes (between participants and nonparticipants) are either insignificant or where there are significant changes they tend to increase the disadvantage of participants (negative effect size). In only one case the over-time gap change is significant and positive, albeit small (0.1 *sd*), and this is the effect of Title I on math for 3rd-graders after four years of Title I exposure. This positive over-time gap change of 0.1 *sd* indicates that by the time 3rd grade participants completed 6th grade, they had gained only about one-tenth of a standard deviation more in math achievement than nonparticipant students.

3.3. The 1999 National Assessment and Follow-up Studies

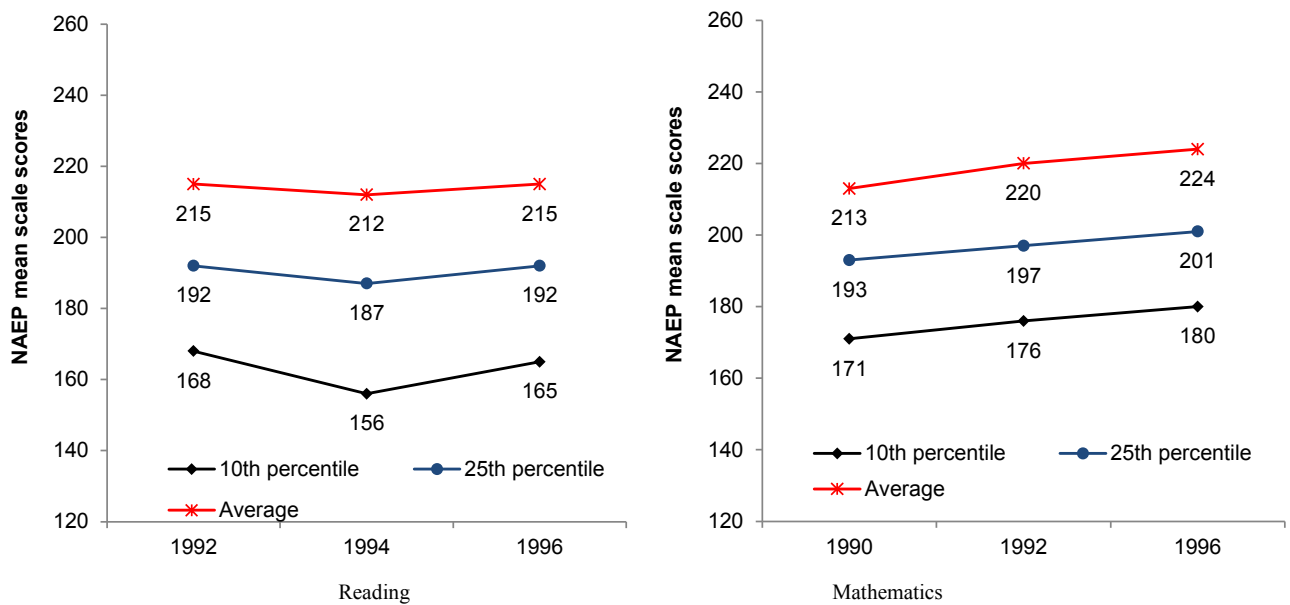
The 1999 National Assessment of Title I was mandated by the Congress as part of the 1994 reauthorization of ESEA [18]. A few follow-up studies conducted between 1999 and 2001 are summarized in [19]. Unlike the Prospects study, the 1999 National Assessment and its follow-up studies do not compare the academic achievement of Title I participants versus nonparticipants. Rather, they conduct a trend analysis of national-level NAEP achievement test scores in reading and math from 1986 to 1999 for subgroups of students from disadvantaged backgrounds and thereby likely to benefit from Title I, as compared to those of non-disadvantaged counterparts. These groups are: (a) 9-year-old students in high-poverty schools (schools were 75 percent or more students receive free- or reduced-price lunches), from 1986 to 1999 and (b) low-achieving (below the 10th percentile) 4th-grade students, from 1992 to 1999. These subgroups are likely to benefit from Title I because Title I is designed to support schools with high concentration of poverty, particularly students in greatest risk of failing, and because most Title I funds serve elementary schools. NAEP scores provide a uniform basis for comparing achievement progress nationwide.

One limitation of this trend analysis approach is that it is hard to claim that achievement progress, even for those students likely to receive Title I, are in fact attributable to Title I. Although Title I is the largest single federal educational program, Title I accounts for only about 3 percent of total resources invested in elementary and secondary education by federal, state, and local authorities combined. However, the expansion of school-wide programs funded by Title I, which started in 1996, blurred the distinction between program participants and nonparticipants making it hard to find approaches that contrast these two groups [18].

¹ Following [17], we classify the magnitude of the effect as: (a) small, if it is below 0.22 *sd* or 10 scale-score points; (b) moderate, if it is between 0.22 *sd* and 0.56 *sd* (or between 10 and 25 scale-score points); and (c) large, if it is over 0.56 *sd* or 25 scale-score points.



A) Average scale scores on the Trend NAEP of 9-year-old public school students, by poverty level



B) Average scale scores on the Main NAEP of 4th-graders public school students, by performance percentile

Figure 3. Performance on NAEP for Students Most Likely to Benefit from Title I Services, 1990 to 1999

The results of the 1999 National Assessment are summarized in Figure 3. Our main outcome measure of interest is the achievement gap between disadvantaged groups, measured in this study as students in low-poverty schools (Figure 3A) and students in the 10th achievement percentile (Figure 3B), and their non-disadvantaged counterparts.

The data required for conducting significance tests and calculating standardized effect sizes is neither reported nor available at the NAEP data portal [24]. Still, achievement score-point trends indicate that during the period from 1988 to 1999, students from disadvantaged backgrounds fail to reduce their achievement gaps vis-à-vis their non-disadvantaged counterparts. Two major findings

support this overall assessment. First, over the decade from late 1980s to late 1990s the achievement gaps widen for both groups of disadvantaged students, as compared to their non-disadvantaged counterparts and it does so in both reading and math. The gap between highest- and lowest-poverty schools for 9-year-old students, as measured by average NAEP scores, widens from a 27-point gap in 1988 to a 40-point gap in 1999 in reading; and from a 20-point gap in 1986 to a 29-point gap in 1999 in math. The performance gap for the lowest-performing 4th-graders as compared to the overall average widens from 47 to 50 points between 1992 and 1996 in reading and from 42 to 44 points from 1990 to 1996 in math. Second, achievement gaps in the late 1990s are substantial, equivalent to several grade levels.

A 10-point difference in NAEP scale scores is considered roughly equivalent to one grade level [19]. In the late 1990s 9-year-old students in the highest-poverty schools have an average gap vis-à-vis the lowest-poverty schools equivalent to four grade levels in reading (40-point gap) and nearly three grade levels in math (29-point gap). Similarly, the lowest-performing 4th-graders have a gap vis-à-vis the overall average equivalent to five grade levels in reading (50-point gap) and four grade levels in math (44-point gap).

3.4. The 2007 National Assessment and Follow-up Studies

The 2007 National Assessment of Title I was mandated by Congress as part of the 2001 NCLB Act and commissioned by the Institute of Education Sciences in the Department of Education [6,7,20]. A set of follow-up evaluation studies conducted since 2007 are summarized in a report released in 2009 and conducted by the Policy and Program Studies Service in the Department of Education [21]. The 2007 National Assessment analyses national-level NAEP trends for groups of disadvantaged students considered as the most likely beneficiaries of Title I under the 2001 NCLB mandate. This methodological approach is similar to that followed by the 1999 National Assessment. The two national assessments vary slightly in the groups of disadvantaged students they consider. Like its predecessor study, the 2007 National Assessment also evaluates students in highest-poverty schools, but unlike its predecessor it includes racial and ethnic minorities, specifically black and Hispanic students, as opposed to low-achieving students.

As with the 1999 study, the major limitation of the 2007 National Assessment is that the achievement trend data do not isolate the impact of Title I. Rather it measures the effect of Title I and the entire educational system on students' achievement. Simple trend analyses such as the ones done in this study cannot separate the effects of Title I from the effects of other state and local improvement programs, demographic changes, and other factors that may affect student achievement trends.

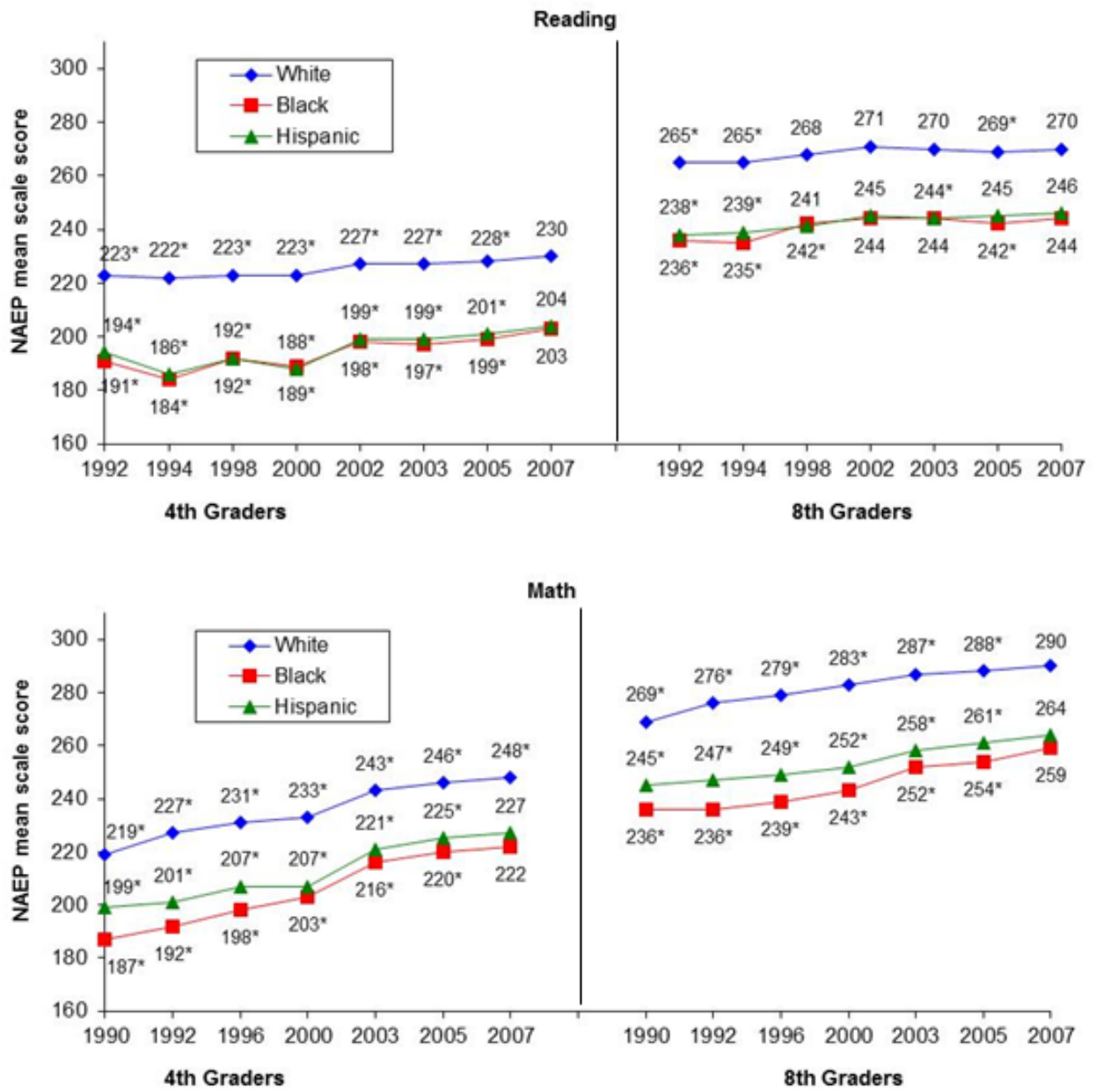
Figures 4 to 6 summarize the national-level trends in NAEP test scores for 4th- and 8th-graders in reading and math, by race/ethnicity and by school poverty level, from 1990 to 2007. We do not include the NAEP scores in science in our analysis because there are only three data points: 1996, 2000 and 2005. We draw on these NAEP trends to assess the effects of Title I on students' achievement according to our outcome measure of interest: the achievement gaps between disadvantaged groups, measured in this case by students in low-poverty schools and students from racial/ethnic minorities, and their non-disadvantaged counterparts. We report gaps in scale-score points, obtained directly from the

figures, and also as effect sizes, using the corresponding standard deviations obtained from the NAEP data portal [24].

The results by race/ethnicity in Figure 4 show that in reading the black-white gap in 4th grade reduces by 5 score points or 0.13 *sd*, from 32 points (0.93 *sd*) in 1992 to 27 points (0.8 *sd*). Most of the improvement occurs in a single interval, between 2000 and 2002. The reduction in the black-white gap in 8th grade reading is smaller, from 29 points (0.89 *sd*) to 26 points (0.8 *sd*); a closure of 3 points or 0.09 *sd*. The change in the Hispanic-white gaps in reading is similar in magnitude to the black-white gap in both grades. In math, there is nearly continuous improvement between 1990 and 2007 for all groups in both grades, although the gains have been greater in 4th grade. The 4th grade math black-white gap narrows by 6 points or 0.12 *sd*, from 32 points (1.1 *sd*) to 26 points (0.98 *sd*), while the 8th grade math gap narrows slightly (2 points; less than 0.01 *sd*) from 33 points (0.965 *sd*) to 31 points (0.964 *sd*). The Hispanic-white math gap widens slightly for 4th graders (by only 1 score point; 0.07 *sd*) and also for 8th graders (by just 2 score points; 0.08 *sd*).

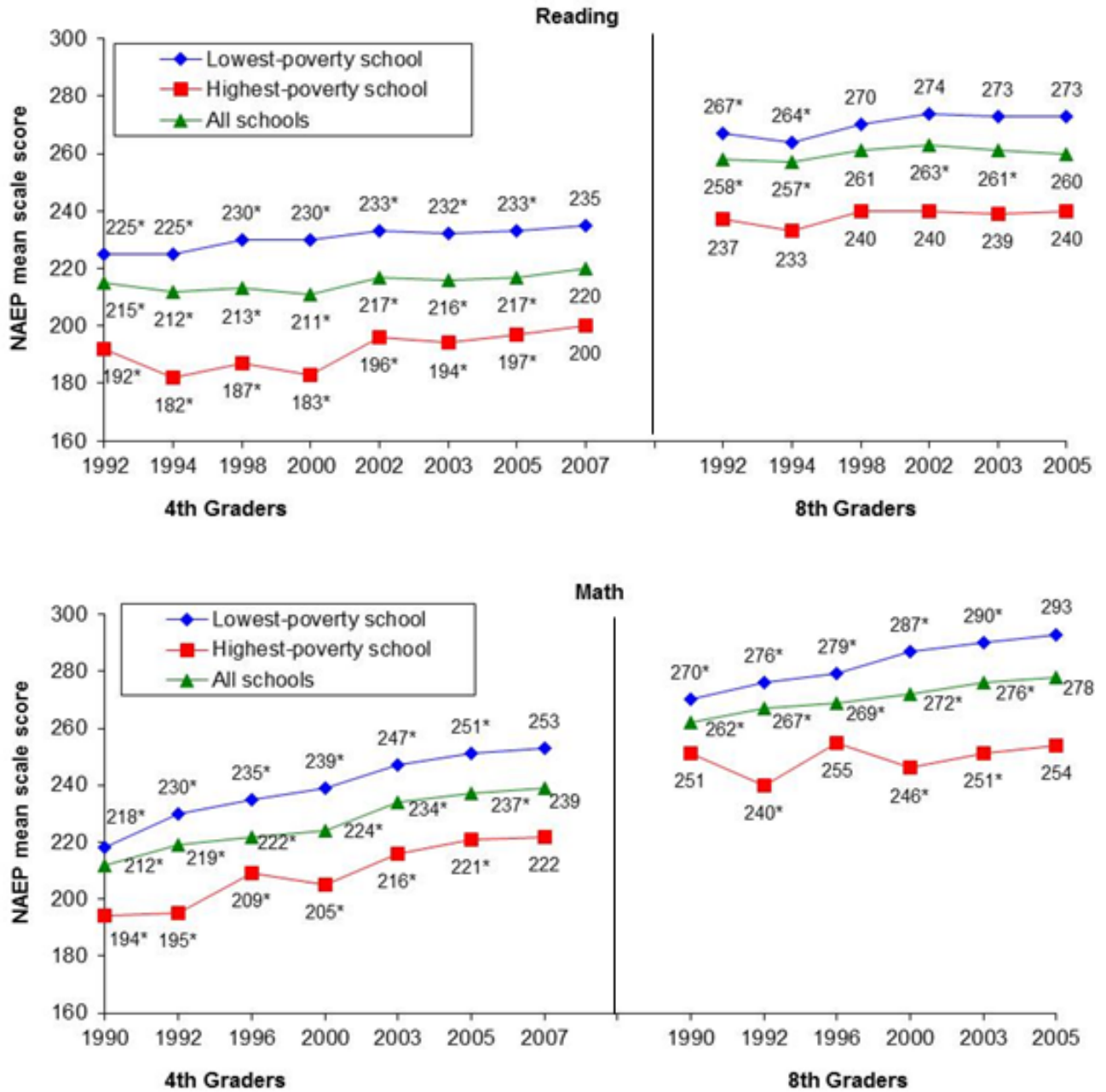
Figure 5 shows an unfavorable pattern when the highest-poverty schools are compared to the lowest-poverty schools. Between 1992 and 2007, the poverty gap in reading widens slightly for both 4th graders (by 2 score points; 0.04 *sd*) and 8th graders (by 3 score points; 0.08 *sd*). In math, the poverty gap widens 7 score points (0.24 *sd*) for 4th graders (from 1990 to 2007) and 20 points (0.59 *sd*) for 8th graders (from 1990 to 2005). Despite the improvement in math scores for both groups and both grades over the 1990 – 2007 period, the gains are greater for low-poverty schools than for high-poverty schools, which results in the widening of the poverty gaps at the school level.

When we consider the percentage of students at or above the NAEP proficiency level between 1992 and 2007, shown in Figure 6, we find that the improvement of white students is greater than both black and Hispanic students in both subjects and in both grades. In reading, between 1992 and 2007, black-white gap widens by 3 percentage points in 4th grade and by one percentage point for 8th graders. The widening of the Hispanic-white reading gaps is similar in magnitude. In math, there is a continuous improvement in the percentage of students performing at or above the NAEP proficiency level between 1990 and 2007, for all groups and in both grades. However, whites improved at a faster rate and as a result the gaps widens substantially. The black-white math gap widens 22 percentage points for 4th graders and 17 percentage points for 8th graders. The Hispanic-white gap widens by 18 percentage points in 4th -grade and by 15 percentage points in 8th grade.



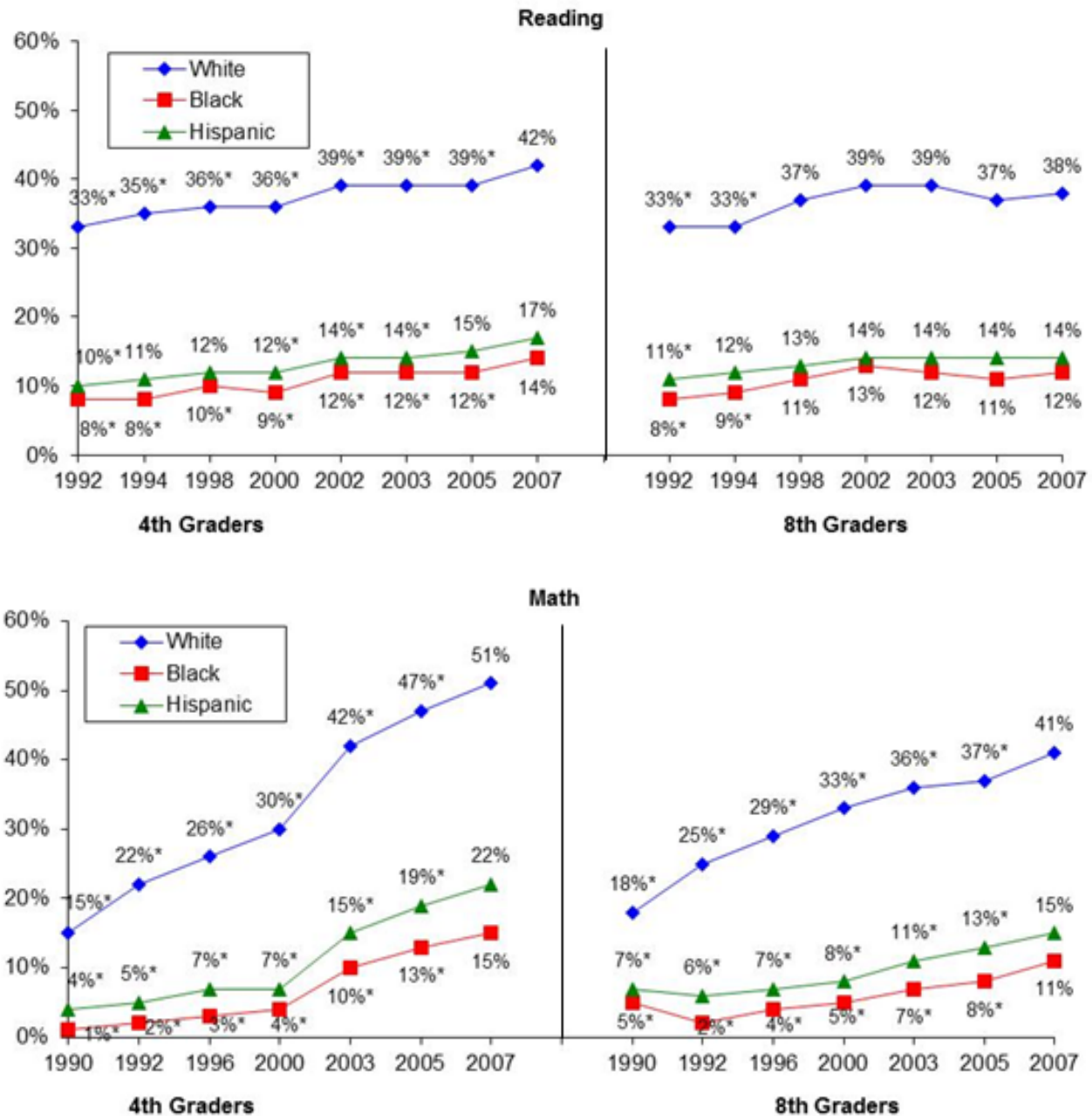
* $p < .05$

Figure 4. Performance on main NAEP for Public School Students by Race and Ethnicity, 1990 to 2007



* $p < .05$

Figure 5. Performance on main NAEP for Public School Students by School Poverty Level, 1990 to 2007



* $p < .05$

Figure 6. Students at or above the NAEP Proficiency Level by Race and Ethnicity, 1990 to 2007

In short, the 2007 National Assessment of Title I suggests that Title I was largely ineffective in closing the achievement gap of students from disadvantaged backgrounds over the period from the early 1990s to the late 2000s. As measured by overall scale scores, there is a modest closure of the black-white and Hispanic-white gaps in the 4th grade over this period, but very little change in the 8th grade gaps. There is no closure in the achievement gaps between high-poverty schools and more affluent schools at either grade over the same period. The black-white gap closure for 4th graders represents a standardized effect of only about 0.13 *sd* in reading and 0.12 *sd* in math, and these are achieved mostly in a single interval – between 2000 and 2002. As measured by proficiency rates, the black-white and Hispanic-white gaps

show little change in reading, but the math gaps worsen substantially. This suggests that the improvements in minority math scores gap in the 4th grade might be mainly at the lowest end of the performance continuum, raising some of the lower scores to some degree but not sufficient to cross the proficiency threshold. Part of these improvements for the lowest-performing students might have been due to special accommodations in testing administrations, which were adopted in the late 1990s and whose use increased markedly between 2000 and 2003.

3.5. The Dee and Jacob Study

The study by Dee and Jacob [22] analyzes state-level

NAEP panel data from 1992 to 2007 to determine whether NCLB has influenced student achievement nationally. They use a comparative interrupted time series (CITS) approach (also known as an interrupted time series with a non-equivalent control group) to contrast states with and without school accountability policies in place prior to NCLB. Their outcome measure is over-time achievement gaps between states without (treatment states) and with (control states) pre-NCLB accountability policies. Specifically, the study compares the deviation from prior trend for the states that were arguably affected by NCLB provisions (those without pre-NCLB accountability policies; the treatment group) with the analogous deviation for states that were less affected by NCLB (those with pre-NCLB accountability policies; the control group). The rationale is that the deviations from prior achievement trends within states with pre-NCLB accountability policies (control states) provide a good counterfactual for what would have happened in states without pre-NCLB policies (treatment states) if NCLB had not been implemented.

One advantage of this study is that the panel-based research design allows distinguishing the effect of NCLB from the effects of other state and local educational changes, as well as other social and economic changes that took place during the period of analysis. This is a limitation of both the 1999 and 2007 National Assessment studies. One limitation of the study by Dee and Jacob is that it has data for only a subset of states ranging from 19 to 39 depending on the subgroup (see Table 2). This raises issues about the representativeness of the results nationally. Some of the states missing from this analysis have very large student (and minority) populations, including Florida, Illinois, New Jersey, and Pennsylvania. The number of states considered in the analysis is particularly small for Hispanics (ranging from

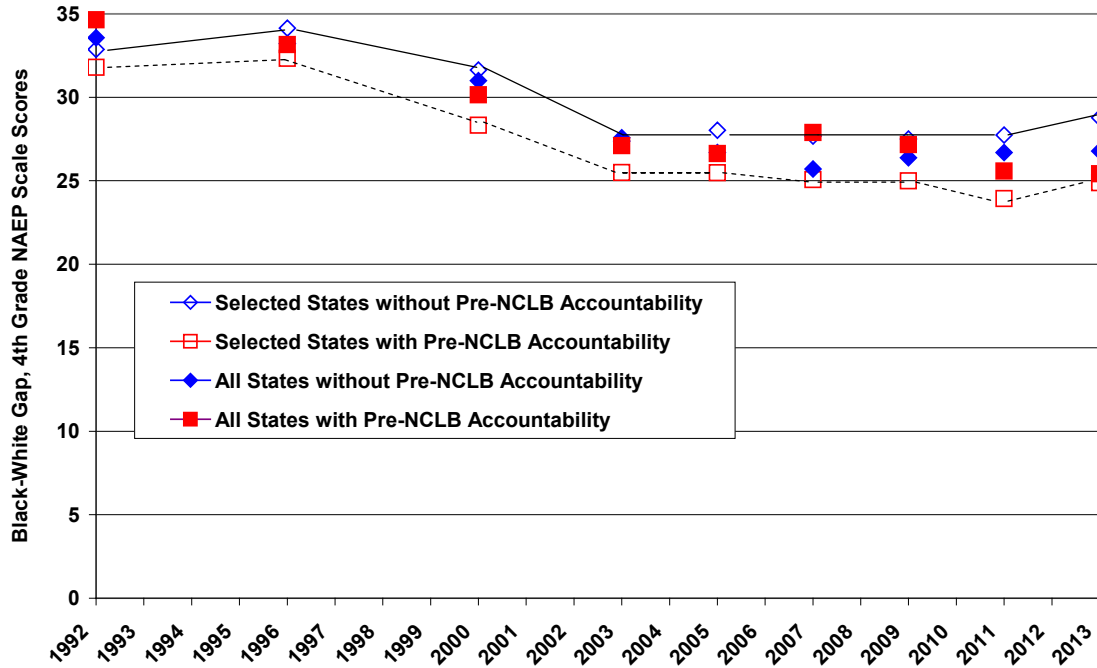
16 to 22 states depending on the grade-subject) and for blacks (ranging from 27 to 32 states), which raises concerns about the representativeness of the results for these subgroups with respect to the nation as a whole. Table 2 summarizes the estimated effects of NCLB on NAEP scores for 4th- and 8th-grade math and 4th-grade reading. The effects are available at the aggregated school level and by race, free-lunch eligibility and proficiency level.

The results at the aggregate school level show that by 2007 NCLB has a moderate positive effect for 4th-grade math (7.2 score points above control states) but smaller and statistically insignificant effects for 8th-grade math (3.7 points) and 4th-grade reading (2.3 points). Since one of the primary objectives of NCLB is to reduce the achievement differences by race and socioeconomic status, we are primarily interested in the effects of Title I/NCLB on subgroups of students from disadvantaged backgrounds and not so much on the effects at the aggregate school level. However, given that the analysis by Dee and Jacob for subgroups of students does not cover the same group of states, we can only interpret them in terms of effects on a particular subgroup but we cannot infer what the results mean with respect to changes in gaps between disadvantaged and non-disadvantaged groups of students (which is our outcome measure of interest). We address this limitation by complementing Dee and Jacob analysis with our own analysis for subgroups of students covering the same group of states (more below). Dee and Jacob results show moderate effects by race/ethnicity and greater for blacks and Hispanics than for whites for 4th-grade math; somewhat smaller effects for 8th-grade math, greater for blacks but of a similar magnitude for whites and Hispanics; and a combination of significant effects for whites but insignificant effects for blacks and Hispanics in 4th-grade reading.

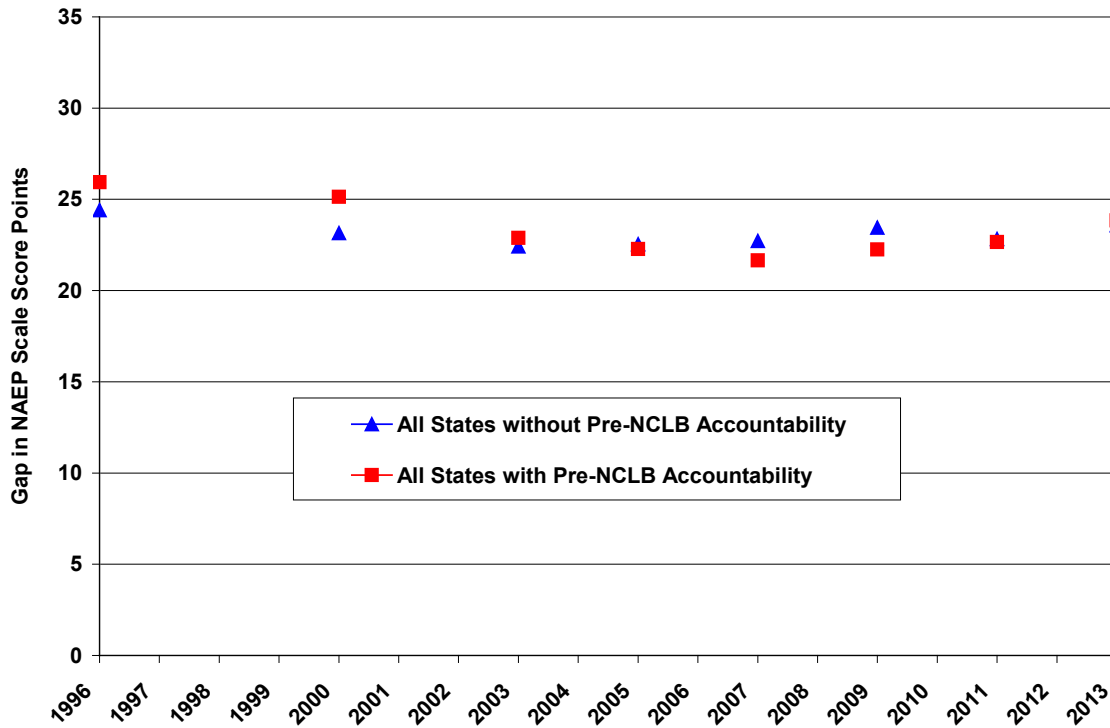
Table 2. Summary of Estimated Effects of NCLB in Dee and Jacob Study

	4th-grade math scores	8th-grade math scores	4th-grade reading scores
Aggregate school level	7.244**	3.704	2.297
Number of states	39	38	37
Race			
White	4.855**	1.828	5.362**
Number of states	39	38	37
Black	14.573**	8.826	-0.871
Number of states	30	27	32
Hispanic	9.793**	8.219**	0.242
Number of states	19	16	22
Free-lunch eligibility			
Eligible	8.011**	15.761**	2.482
Not eligible	1.385	0.992	-4.79
Number of states	36	34	37
Proficiency level			
10th percentile	9.046**	5.598**	3.611
90th percentile	5.205**	2.537	2.097**
Number of states	39	38	37

* $p < .05$. ** $p < .01$.



A) Black-White gap, Grade 4 math



B) Free lunch eligibility gap, Grade 4 math

Figure 7. Gaps for grade 4 math by pre-NCLB Accountability Status.

Being particularly interested in the gap between whites and other minority groups, we conducted a further analysis of NAEP scores showing changes in the 4th grade black-white gap in math, comparing the group of 39 states in the Dee and Jacob study with all 50 states. These results are shown in Figure 7, where “selected states” are the 39 states with NAEP testing in 2000 and at least two scores between 1992 and 2000. The “All States” group includes the 11 states

not used in the Dee & Jacob analysis, which include Colorado, Florida, Illinois, New Jersey, Pennsylvania, Washington, Oregon (black scores only) and four other states. “Free lunch eligibility gap” is the gap between those eligible and not eligible for free/reduced lunch meals.

Both the 39 selected states in the Dee and Jacob study and the total population of 50 states show that the black-white achievement gap in 4th grade math declined appreciably

between 1996 and 2003, a time period that encompasses most of the state and federal policy changes in accountability (Figure 7A). The 4th grade math gap stood at about 33 points in 1996, and fell in two steps to about 27 points by 2003, a standardized effect size of about 0.2 *sd*. We note that the accountability states falling more sharply between 1996 and 2000 than the non-accountability states. However, the gaps changed very little after 2003, dropping to about 25 points and remaining there until 2013. States without pre-NCLB accountability drift upwards slightly, particularly those in the selected states without scores in the year 2000. There are some notable differences in the gaps when comparing the 39 states selected by Dee and Jacob versus all 50 states, particularly when looking at the most recent NAEP scores in 2011 and 2013 (Dee and Jacob examined NAEP scores through 2007). The selected samples of states without pre-NCLB accountability trend upwards in 2013, while all states remain relatively flat. Using data from all states, the gap in the no-accountability states is only two points higher than that in the accountability states (27 vs. 25); just one point higher than it was in 1992 (34 vs. 33). Thus NCLB appears to have reduced the black-white achievement gap during the early years of implementation, but the gap has remained stubbornly stable for the past 10 years or so.

Looking at Dee and Jacob results by poverty subgroups, the authors found a major effect of accountability for students in poverty, as indicated by free lunch status. Again, only 36 states were used for this analysis, because there were only two NAEP assessments before 2003 with information on free lunch eligibility (1996 and 2000), and 14 states were missing scores for one or both of these assessments (see Table 2). However, our own analysis for all states by pre-NCLB accountability status (Figure 7B) shows that the missing states made less of a difference on the poverty gap than it did for the black-white gap; no difference was greater than a single scale score point. Figure 7B shows the 4th grade math gap between those eligible for free/reduced lunch and those not eligible, and the gap is compared for all states without pre-NCLB accountability policies to all states with accountability. The reduction in the gap attributed to NCLB is more modest than in the black-white gap, and it is just slightly greater for the accountability states. The gap for the accountability states was 27 points in 1996, falling to 22 points by 2003, versus a reduction from 24 to 22 for the non-accountability states.² Like the black-white gap, the eligible vs. non-eligible gap remained constant for the next 10 years.

In short, the study by Dee and Jacob suggests that the NCLB policies reduced achievement gaps modestly for 4th grade math but much less for 8th grade math and reading. Perhaps most important, the NCLB has definitely not met its objective of eliminating achievement gaps between white and minority students, even though both white and minority students have experienced increasing achievement levels in math.

3.6. Supplemental Analysis: NAEP Trends Until 2013

We supplement the information from the major evaluation studies with additional trend analyses of NAEP test scores from 1992 to 2013 to bring the achievement trends shown in the 2007 National Assessment up to date. For the poverty trends, we shift to students' poverty status rather than school poverty status (Figure 8). We also update the trend data by race/ethnicity groups, which are shown Figure 9. We used data from [24] in both Figure 8 and Figure 9.

Figure 8 shows the trends in NAEP reading and math scores by students' individual poverty status. For reading, between 1998 and 2013 4th grade scores rise modestly by 10 points for non-poverty students in a nearly linear fashion, and they also rise by 9 points for free/reduced lunch students but most of the gain was in the first half of the period. Thus the reading gap between poverty and non-poverty 4th graders has remained constant for 15 years. Reading also shows overall improvement for 8th graders, although it is slightly less at 9 points for non-poverty students and 8 points for free/reduced students. The pattern differs, somewhat, in that scores were relatively flat for both groups until 2007, so most of the increases for 8th grade reading have come between 2007 and 2013. Like 4th graders, the reading gap for 8th graders has not closed over this 15 year period. For math, the gains have been much greater, and they continue rising for both groups and both grades after 2007. For 4th graders, the total increase between 1996 and 2013 is 23 points for both poverty and non-poverty students, but most of the increase occurred between 1996 and 2007, particularly for the poverty students. NCLB might have been responsible for the especially steep increase of 12 points for poverty students between 2000 and 2003. For 8th graders, math scores rise 18 points for both groups in a nearly linear fashion. Like 4th graders, the math gap between poverty and non-poverty students remains constant over this 17 year period.

2 There are no NAEP math scores by free/reduced lunch status prior to 1996, so it is possible that the gap was larger in the early 1990's. Additionally, there are no reading scores prior to 1998.

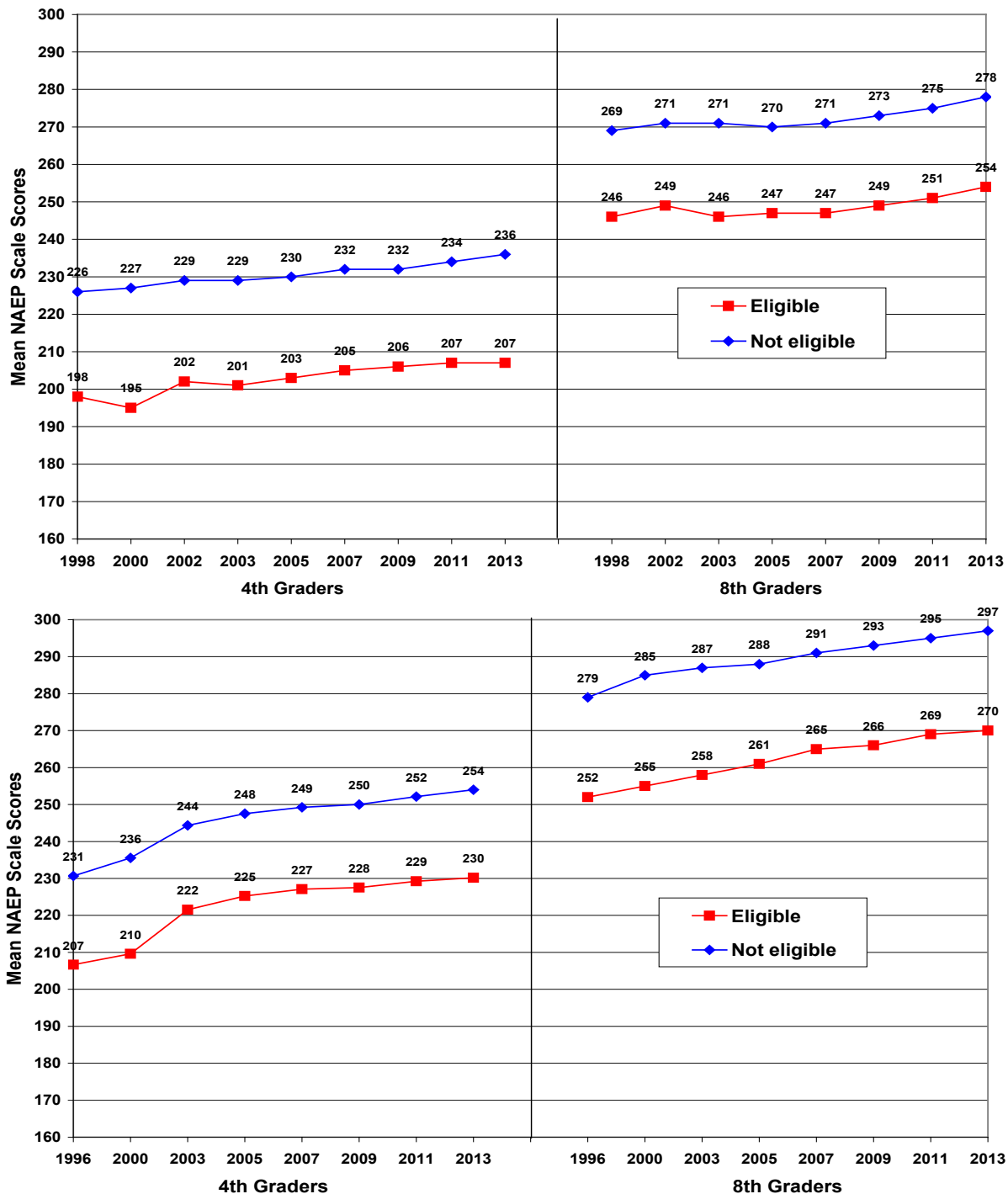
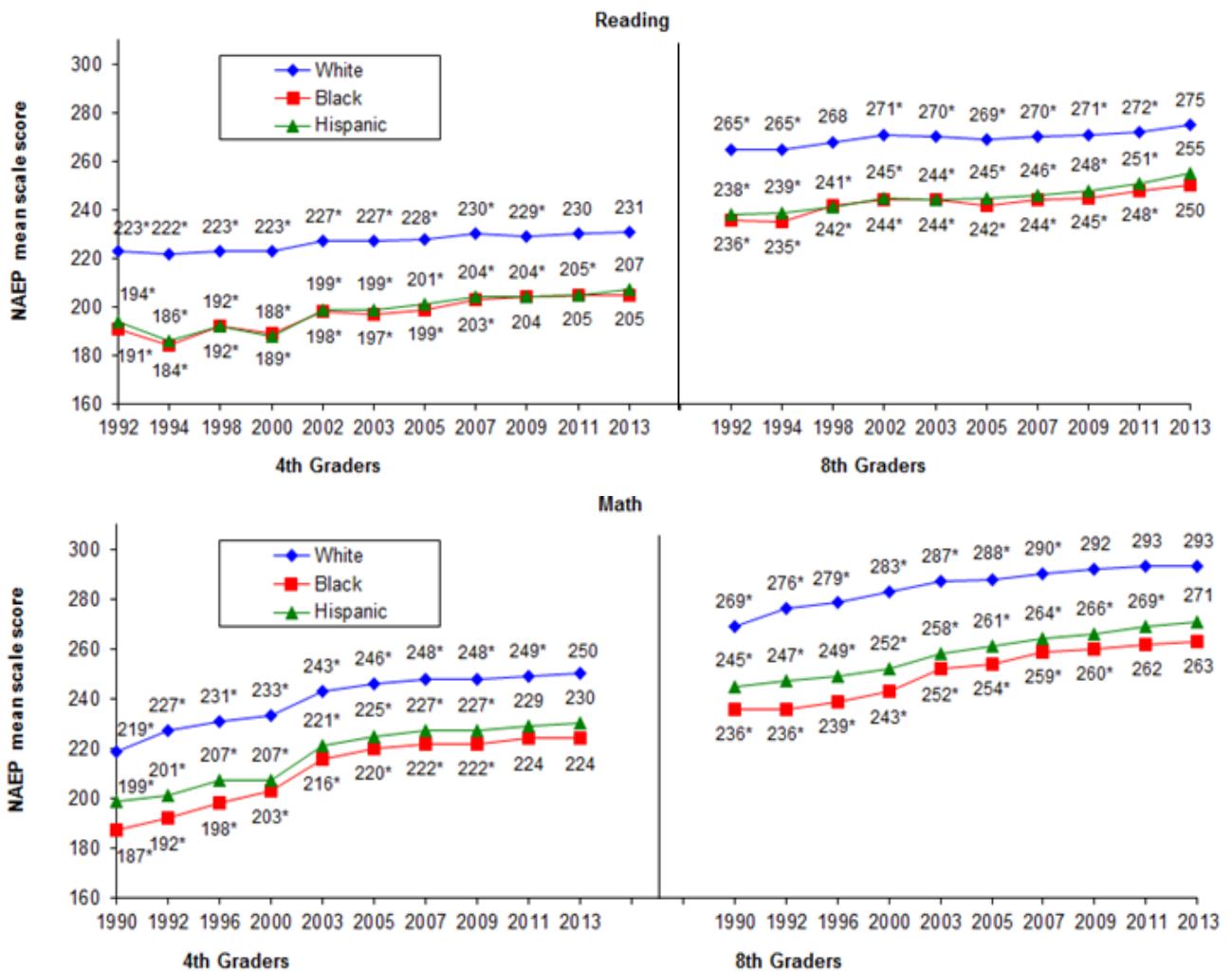


Figure 8. NAEP Reading and Math Scores for Public School Students by Poverty Status (Free Lunch Eligibility), 1996 to 2013

Figure 9 shows reading and math trends by race and ethnicity. Not surprisingly, like the poverty trends these results also reveal a generally upward trend for all groups starting in 1992 for reading and 1990 for math. For 4th grade reading, the trend for white students is flatter than for black or Hispanic students, increasing just 8 points over the 21 year period – less than half a point per year. This compares to an increase of 14 points for black students and 13 points for Hispanics. So the 4th grade black-white reading gap reduced by 6 points for black students (or 0.2 *sd*) and the Hispanic-white gap reduced by 5 points (or 0.02 *sd*) over this

21-year period. The trends for 8th grade reading are similar, with a closure of the black – white gap of 4 points (0.1 *sd*) and a closure of the Hispanic-white gap of 7 points (0.2 *sd*). The growth in math scores has been greater for all groups in both grades, with the 4th grade gains being larger than 8th grade gains (30+ points vs. 20+ points). Black students at both grade levels gain somewhat more than white students, so the black-white gap is reduced by 6 points for 4th graders (0.1 *sd*) but just 2 points (0.05 *sd*) for 8th graders. The Hispanic-white gap remains unchanged for Hispanic 4th graders and reduced by 2 points (0.03 *sd*) for 8th graders.



* $p < .05$

Figure 9. NAEP Reading and Math Scores for Public School Students by Race/ethnicity, 1996 to 2013

The overall impression given by the long term trends in NAEP scores is that there is an improvement in achievement for all grades, groups, and subject matters, but with greater gains in math than reading and somewhat greater gains for black and Hispanic students than white students. This suggests that Title I and NCLB reforms have succeeded more in raising achievement levels for all students than in closing achievement gaps for disadvantaged children.

4. Conclusions

The review of multiple national-level evaluations of Title I/NCLB covering the period from 1966 to 2013 offers very little evidence that the Title I compensatory education program has significantly improved the academic achievement of disadvantaged students nationwide. The earliest national study, the meta analysis of Borman and D'Agostino covering studies between 1966 and 1993, did show modest Title I effects on math during elementary

grades (about 0.2 *sd*) but much lower effects in higher grades and for reading in all grades (0.1 *sd*). The Prospects study covering 1991 to 1994 showed no significant reduction of the achievement gaps between Title I participants and nonparticipants. The final early study, covering the period 1988 to 1999, was an evaluation carried out by the Department of Education which compared 4th grade reading and math scores for highest poverty schools to lowest poverty schools. Generally, the achievement gap widened between lowest and highest poverty schools over this period.

The lack of meaningful gap reductions in the evaluations undertaken before 2000 is contrasted somewhat by evidence provided in later studies, including an evaluation by the U.S. Department of Education in 2007 and a study by Dee and Jacob in 2011. Both of these studies suggest that No Child Left Behind had modest effects on 4th grade test scores, especially in math, and these gains were somewhat stronger for disadvantaged students. According to the Department of Education study, the gains for disadvantaged students and schools took place primarily between 2000 and 2002, which

corresponds to the implementation of NCLB.

Stronger evidence that NCLB accountability improved the achievement of disadvantaged students was provided by the Dee and Jacob study, which used a quasi-experimental method to compare states that had NCLB-type accountability reforms prior to the national law. Focusing on the achievement gap per se, the authors conducted a further analysis of the states selected for the Dee & Jacob study. This analysis showed that states with pre-NCLB accountability reduced the black-white gap 7 points by 2003 compared to 5 points for states without accountability.

In order to summarize the progress made on closing achievement gaps, a final analysis was carried out by the authors using NAEP data that covers 1990 to 2013, more than two decades during which various Title I programs and policies were in place. The overall progress is disappointing, particularly for the poverty gap. The achievement gaps between students eligible for free/reduced lunch vs those not eligible have remained virtually constant for reading and math at both grade levels. The picture is more positive for black-white and Hispanic-white gaps, particularly in 4th grade. At that grade level, both of these gaps have been reduced by about 6 scale score points, which is a standardized effect size of slightly less than 0.2 *sd*. Reduction in the 8th grade reading and math black-white gaps are only 4 and 3 points, respectively.

The different progress in gap reductions before and after 2000 could reflect the very different policy approaches of these periods. The original concept behind Title I was to establish compensatory programs for disadvantages students, on the assumption that extra remedial instruction would allow these children to catch up. In retrospect, it might be fairly argued that the level of funding, which rarely exceeded \$1500 per student (in 2012 dollars), could not be expected to close achievement gaps, given the difficulty of this task.

The national approach after 2000 was quite different, it was to adopt accountability practices which had proven effective in some states during the late 1990s. Thus NCLB was a systemic reform that aimed to raise achievement by standardizing curriculum, adopting uniform standards, and publishing results by demographic group, all at the state level. It is not completely clear why policymakers assumed this would raise achievement levels of disadvantages students rather than raise achievement for everyone; at least that has not been clearly articulated. To some extent, this same thinking is behind the most recent attempt at standards reform, which is to adopt a common core of curriculum and standards that would be adopted nationwide. Based on the failure of No Child Left Behind to close achievement gaps, it is unlikely that Common Core will do so either, whatever it does for overall achievement levels.

REFERENCES

[1] Education Department Budget History, Online available from

<http://www2.ed.gov/about/overview/budget/history/index.html>

- [2] ED Data Express, Online available from <http://eddataexpress.ed.gov/index.cfm>
- [3] H. Cooper, L. V. Hedges, J. C. Valentine (eds.). The handbook of research synthesis and meta-analysis. Russel Sage Foundation, 2009.
- [4] S. W Raudenbush. New directions in the Evaluation of Title I. American Educational Research Association annual meeting, New Orleans, 2002.
- [5] U.S. Department of Education. Final report on the national assessment of Title I: Summary of key findings, U.S. Government Printing Office, Washington. DC, 2007.
- [6] J. Torgesen, A. Schirm, L. Castner, S. Vartivarian, W. Myers, F. Stancavage, D. Durno, R. Javorsky and C. Haan. National Assessment of Title I, Final Report: Volume II: Closing the Reading Gap, Findings from a Randomized Trial of Four Reading Interventions for Striving Readers (NCEE 2008-4013), U.S. Government Printing Office, Washington. DC, 2007.
- [7] U.S. Department of Education. State and Local Implementation of the No Child Left Behind Act Volume I—Title I School Choice, Supplemental Educational Services, and Student Achievement, U.S. Government Printing Office, Washington. DC, 2007.
- [8] U.S. Department of Education. State and Local Implementation of the No Child Left Behind Act, Volume IV—Title I School Choice and Supplemental Educational Services: Interim Report, U.S. Government Printing Office, Washington. DC, 2008.
- [9] U.S. Department of Education. State and Local Implementation of the No Child Left Behind Act, Volume VIII—Teacher Quality Under NCLB: Final Report, U.S. Government Printing Office, Washington. DC, 2009.
- [10] U.S. Department of Education. State and Local Implementation of the No Child Left Behind Act, Volume IX—Accountability Under NCLB: Final Report, U.S. Government Printing Office, Washington. DC, 2010.
- [11] E. Hanushek, M. E. Raymond. Does School Accountability Lead to Improved Student Performance?, *Journal of Policy Analysis and Management*, Vol.24, No.2, 297-327, 2005.
- [12] J. D. Matsudaira, A. Hosek, E. Walsh. An integrated assessment of the effects of Title I on school behavior, resources, and student achievement, *Economics of Education Review*, Vol.31, 1-14, 2012.
- [13] W. Van der Klaauw, Breaking the link between poverty and low student achievement: and evaluation of Title I, *Journal of Econometrics*, Vol.142, 731-756, 2008.
- [14] E. U. Cascio, N. E. Gordon, S.J. Reber. Federal aid and equality of educational opportunity: evidence from the introduction of title I in the south, NBER Working Paper Series No. 17155, National Bureau of Economic Research, 2011.
- [15] M. Wong, T. D. Cook, P. M. Steiner. No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series, Institute for Policy Research, Documento de trabajo 09-11, No.15, 2009.

- [16] G. D. Borman, J. V. D'Agostino. Title I and student achievement: a meta-analysis of Federal evaluation results, *Educational Evaluation and Policy Analysis*, Vol.18, 309–326, 1996.
- [17] M. J. Puma, N. Karweit, C. Price, A. Ricciuti, W. Thompson, M. Vaden-Kiernan. *Prospects: student outcomes*, Final report. Abt Associates, Inc., Cambridge, MA, 1997.
- [18] U.S. Department of Education. *Promising results, continuing challenges: The final report of the national assessment of Title I*. U.S. Government Printing Office, Washington, DC, 1999.
- [19] U.S. Department of Education. *High standards for all students: A report from the national assessment of Title I on progress and challenges since the 1994 Reauthorization*, Government Printing Office, Washington, DC, 2001.
- [20] S. Stullich, E. Eisner, J. McCrary, *Policy and Program Studies Service*. National assessment of Title I. Final report. Volume I: Implementation, U.S. Government Printing Office, Washington, DC, 2007.
- [21] U.S. Department of Education. *Title I Implementation: Update on recent evaluation findings*, U.S. Government Printing Office, Washington, DC, 2009.
- [22] T. Dee, B. Jacob. The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, Vol.30, 418–446, 2011.
- [23] E. McDill, E. G. Natriello. The effectiveness of the Title I compensatory education Program: 1965 – 1997, *Journal of Education for Students Placed at Risk*, Vol.3, 317–335, 1998.
- [24] Online available from:
<http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx>