Does Money Matter for Minority and Disadvantaged Students? Assessing the New Empirical Evidence

David Grissmer

Ann Flanagan

Stephanie Williamson

RAND

About the Authors

David Grissmer is a senior management scientist at RAND. His education research has included work on teacher supply and demand and the analysis of national and regional test score trends. He is currently engaged in research analyzing state achievement scores, estimating the number and location of children at educational risk in the United States and assessing the federal research and development portfolio on children.

Ann Flanagan is a consultant at RAND. She is analyzing state test scores to determine the effects of family characteristics and educational policies and resources. Her previous research has included estimating the amount and kind of federal research and development on children as well as research on the incidence, effects, and prevention of drug use.

Stephanie Williamson is a research assistant at RAND. Her research areas have included analysis of national and regional test score trends and examination of the relationship between family characteristics and student academic achievement. Her work also includes analyzing reasons for the convergence of the black-white score gap. She is currently investigating academic skills at work, including the exploration of the relationship between various pregraduation experiences and a range of post high school outcomes.

Does Money Matter for Minority and Disadvantaged Students? Assessing the New Empirical Evidence

David Grissmer

Ann Flanagan

Stephanie Williamson

RAND

Until relatively recently, the consensus among social scientists was that providing schools additional resources would have little impact on student achievement—the so-called "money doesn't matter" thesis (Ladd, 1996). This counter-intuitive view actually dates from the "Coleman Report" which found family influence strong and little effects of school resources (Coleman et al., 1966). Influential reviews by Eric Hanushek (1989, 1994, 1996) also argued that evidence from over 300 empirical studies provided no consistent evidence that increased school resources raised achievement scores. While this view was consistently challenged by many educators, policymakers, and parts of the research community, the empirical evidence simply suggested otherwise.

This scholarly consensus began to crack in the early 1990s. Hedges and his colleagues conducted a formal meta-analysis of the studies that Hanushek had reviewed. They found that most of these studies lacked the statistical power to detect resource effects even when they were quite large. When Hedges and his colleagues pooled data from all available studies, the results indicated a positive, statistically significant effect and provided evidence that some programs may have large effects (Hedges et al., 1992; Hedges and Greenwald, 1996). Other work conducted with alternate methodologies like Hierarchial Linear Modeling rather than the "production function" framework used in the econometric community often showed positive effects of resources.[1]

Nevertheless, Hanushek made one argument that was hard to rebut. Measured in constant dollars, per-pupil expenditures (PPEs) doubled between the late 1960s and the early 1990s. Yet the National Assessment of Educational Progress Tests (NAEP) of representative samples of 9-, 13-, and 17-year-old children seemed to show little improvement during the period when resources rose so rapidly. The increases in reading and mathematics scores from the early 1970s to 1992 were between 0.1 and 0.2 standard deviation or about 4–5 percentile points.

However, accumulating evidence is now challenging both the NAEP evidence and the accuracy of previous empirical studies. The accumulative evidence is certainly sufficient to replace the "money doesn't matter" hypothesis with one that states that additional money matters for students from less advantaged backgrounds and minority students, but may not matter for students from more highly advantaged backgrounds. Several lines of research are converging toward this hypothesis. They include the following:

[1] For two recent examples see Gamoran, 1996 and Raudenbush, forthcoming.

- Re-analysis of experimental data on effect of class size
- Evidence that model specifications used in many previous studies involving non-experimental data have been flawed
- Evidence that from 1967–91, increases in available educational resources aimed at increasing regular students' achievement has been markedly overestimated
- Evidence that the more limited real resources available to increase achievement scores from the late 1960s to the early 1990s was disproportionately targeted at minority and lower income children
- Evidence that minority and less advantaged children made substantial gains in test scores in the 1970 to 1990 period, but more advantaged white students made only small gains
- Evidence that the timing of score gains of minority children seem to be related to both the civil rights and war on poverty efforts as well as declines in class size.

A more consistent set of evidence is now emerging which shows that disadvantaged students received the largest resource gains and that large score gains occurred among these students. We first discuss the evidence from NAEP scores and the companion findings concerning resource growth and targeting. We then discuss several hypothesis for large score gains among blacks in the 1970s and 1980s, and the correspondence with experimental data on the effects on class size. Finally we discuss why estimates on the effects of resources from non-experimental data are now being seriously challenged, and probably have to be discounted in favor of the experimental data.

. . . evidence is now emerging which shows that disadvantaged students received the largest resource gains and that large score gains occurred among these students.

Rising Resources and Rising NAEP Scores

The often-quoted evidence that real per-pupil resources doubled in education from the late 1960s to early 1990s while NAEP scores stagnated is flawed on four accounts. First, although mean NAEP scores did not rise much, this was partly because of rapid growth in the low-scoring Hispanic population. When disaggregated, scores for all racial-ethnic groups rose in reading and mathematics for all age groups. Non-Hispanic whites scores rose by smaller amounts, while scores for Hispanic and blacks rose dramatically. Second, the real increase in educational expenditures was far less than the CPI adjusted PPE data would indicate. Use of more appropriate indices for adjustment of educational expenditures due to their labor intensity provides much smaller estimates of real growth. (Rothstein and Miles, 1995; Ladd, 1996a) Third, a significant part of the smaller estimated increase went for students with learning disabilities, many of whom are not tested.[2] A significant part also went for other socially desirable objectives that are only indirectly related to academic achievement. Taking into account better cost indices and including only spending which would have been directed at increasing achievement scores, Rothstein and Miles (1995) concluded that the real increase in per pupil spending on regular students was closer to 30 percent than to 100.

Finally, the association of additional resources with increased test scores depends upon the distribution of the increased spending. The evidence in

[2] All sides agree that a disproportionate fraction of the expenditure increase during the NAEP period was directed toward special education (Lankford and Wyckoff, 1996; Hanushek and Rivkin, 1997). Hanushek and Rivkin estimate that about a third of the increase between 1980 and 1990 was related to special education. NAEP typically excludes about 5 percent of students who have serious learning disabilities. However, special education counts increased from about 8 percent of all students in 1976–77 to about 12 percent in 1993–94. These figures imply that 7 percent of students taking the NAEP tests were receiving special education resources in 1994, compared to 3 percent in 1976–77. This percentage is too small to have much effect on NAEP trends, but it should in principle have had some positive effect.

Rothstein and Miles (1995) shows that a disproportionate amount of resources was directed toward minority and lower income students.[3] Scores of minority students and lower scoring white students all showed large gains. The argument that additional resources did not matter is not applicable to these students. However, if significant additional resources were also directed toward advantaged students, the evidence would show much smaller gains, and the argument that "money doesn't matter" may apply to these students.

NAEP Data[4]

Trends. Figure 1 shows how black and white 17-year-olds' reading and mathematics scores changed between 1971 and 1996.[5] Figures 2 and 3 show the same data for 9- and 13-year-olds. Each score by race is relative to the earliest test score recorded, so a difference between black and white scores at a given year represents a change in the black-white score gap. The following points stand out:

- The black-white gap narrowed for all ages in both subjects due to substantial gains in black students' scores while white students registered smaller gains.
- The black-white gap narrowed the most for 13- and 17-year-olds due to dramatic increases in black scores from the late 1970s to the late 1980s when black gains were 0.6 to 0.7 standard deviation above white gains.

For 13- and 17-year-olds, the gap stabilized or widened in the 1990s due to significant declines in black reading scores and stable black mathematics scores, while white mathematics scores were increasing. By 1996 black students' gains were between 0.2 and 0.6 standard deviations greater than white students' gains. The black-white gap for 9-year-olds narrowed by 0.25 to 0.35 standard deviation by 1996. The pattern of gains among this group is quite different from that of black adolescents, and the pattern also differs somewhat for reading and mathematics. Black 9-year-olds gained more than older blacks during the 1970s and gained less than older blacks during the 1980s. Although reading scores among black 9-year-olds show declines after 1988, unlike adolescent reading scores, they have returned to 1988 levels. Additionally, mathematics scores continued increasing after 1988 among this cohort.

It is important to stress that even when black gains were largest, they never came close to eliminating the black-white gap. The largest reduction in the gap was for 17-year-olds' reading scores between 1971 and 1988. In 1971 the median black

[3] Rothstein's and Miles' data analyzed detailed data in only nine school districts. More national evidence is needed concerning the relative allocation of additional resources among different types of students. There is little doubt that many of the new programs which were initiated or expanded were directed toward minority or low-income children. These included compensatory education programs such as Title 1 and HEADSTART, efforts within states to change to more equitable funding formulas and desegregation initiatives. However, funding also may have increased for advantaged students. More direct evidence is needed from school-district-level analysis of funding trends for high and low income districts.

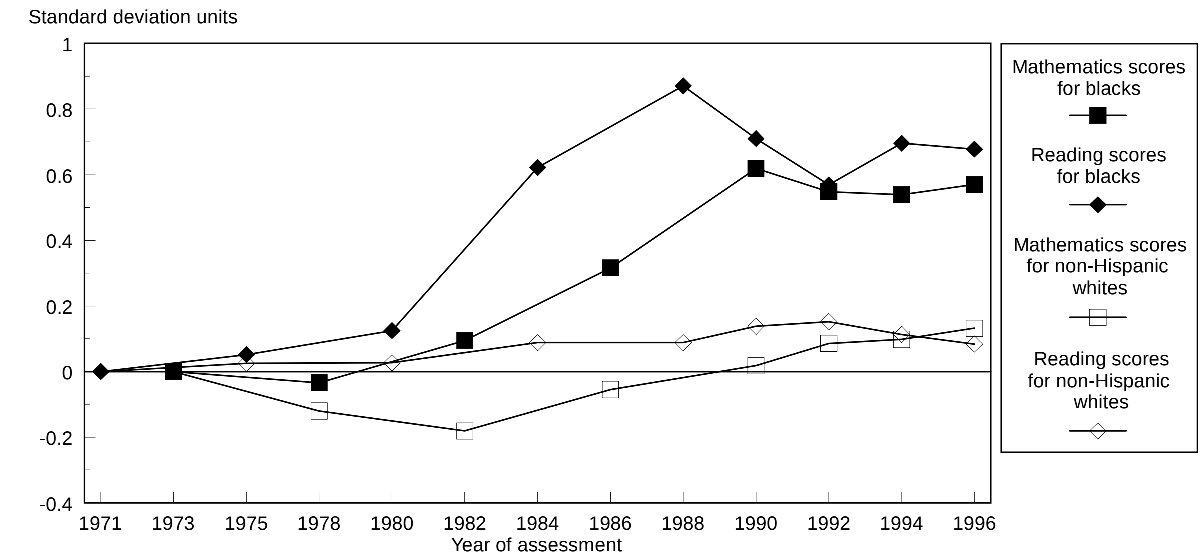
[4] See Cambell et al., 1994 and Cambell et al., 1996 for descriptions of the NAEP data and further references.

[5] The scores have been converted to relative scores by assuming the earliest test score for each race is zero. Thus, the difference in scores reflects changes in the black-white gap from the earliest test. The scores are converted to standard deviation units by taking the mean score difference from the earliest test and dividing by a metric that remains constant over the period—the standard deviations of all students for the earliest year. Another common practice is to measure the gap with respect to the standard deviation in the same year. Since the standard deviation for all students declines for mathematics scores, but increases for reading scores this method changes the metric over time and would result in a somewhat different measure of gap reduction.

The 1973 and 1971 scores for non-Hispanic white students were estimated because the only published scores are for combined Hispanic and non-Hispanic white students in those years. Tests after 1973 have separate data for Hispanic and non-Hispanic white students. We make a small correction in the 1971 and 1973 white data by determining the proportion by age group of students who were Hispanic and assuming that the difference between Hispanic and non-Hispanic white scores were the same in 1971 and 1973 as for the 1975 reading and 1978 mathematics tests.

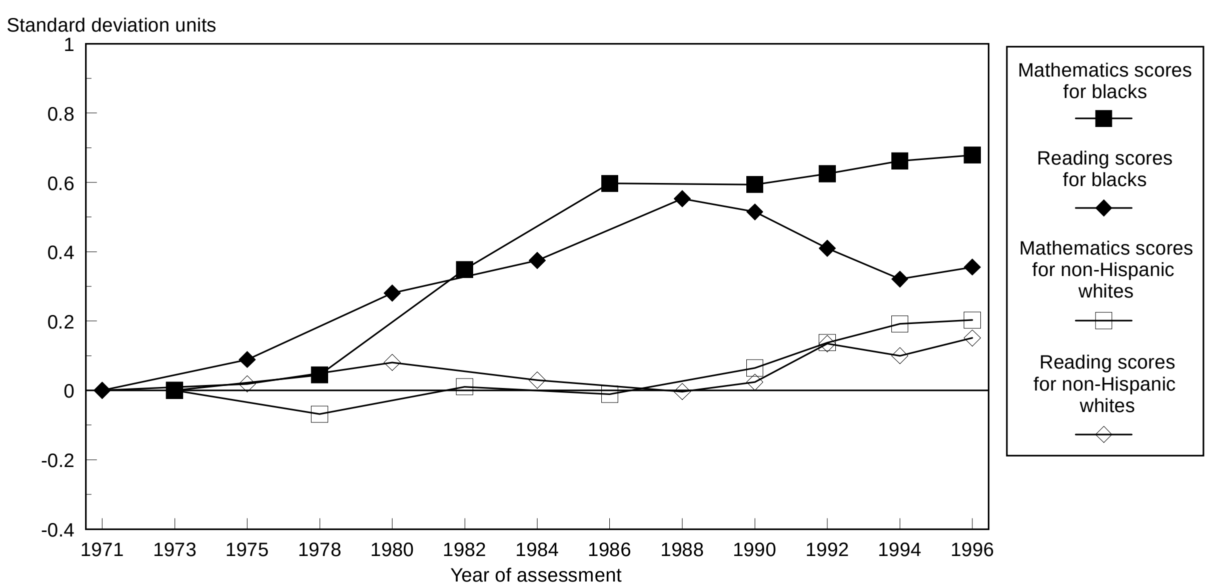
Also the 17-year-old NAEP scores reflect only students rather than all 17-year-olds. Consequently, the 17-year-old scores will be biased with respect to 9- and 13-year-old scores. We make a correction for 17-year-old scores using the proportion of 17-year-olds in school by race in 1971–73 and 1996. We assume that those not tested would have scored one-half standard deviation below the mean score for their respective race—probably a conservative assumption. School enrollment data from the October Current Population Survey (CPS) shows approximately 88 percent of white and 83 percent of black 17-year-olds were in school in 1970 versus 89 and 90 percent in 1988 (Cook and Evans, forthcoming).

Figure 1.—NAEP mathematics and reading scores for 17-year-old students, by race/ ethnicity



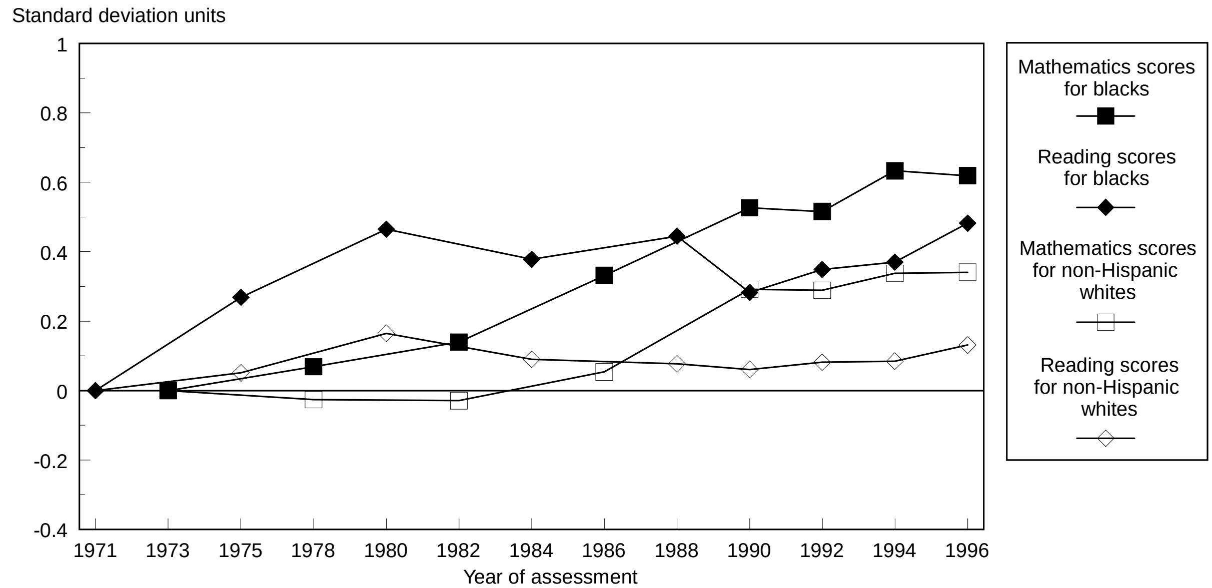
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 2.—NAEP mathematics and reading scores for 13-year-old students, by race/ ethnicity



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 3.—NAEP mathematics and reading scores for 9-year-old students, by race/ethnicity



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

scored between the 10th and 12 percentiles of the white distribution. By 1988 the median black scored between the 26th and 28th percentiles of the white distribution. For the other age groups, the gap remained even wider.

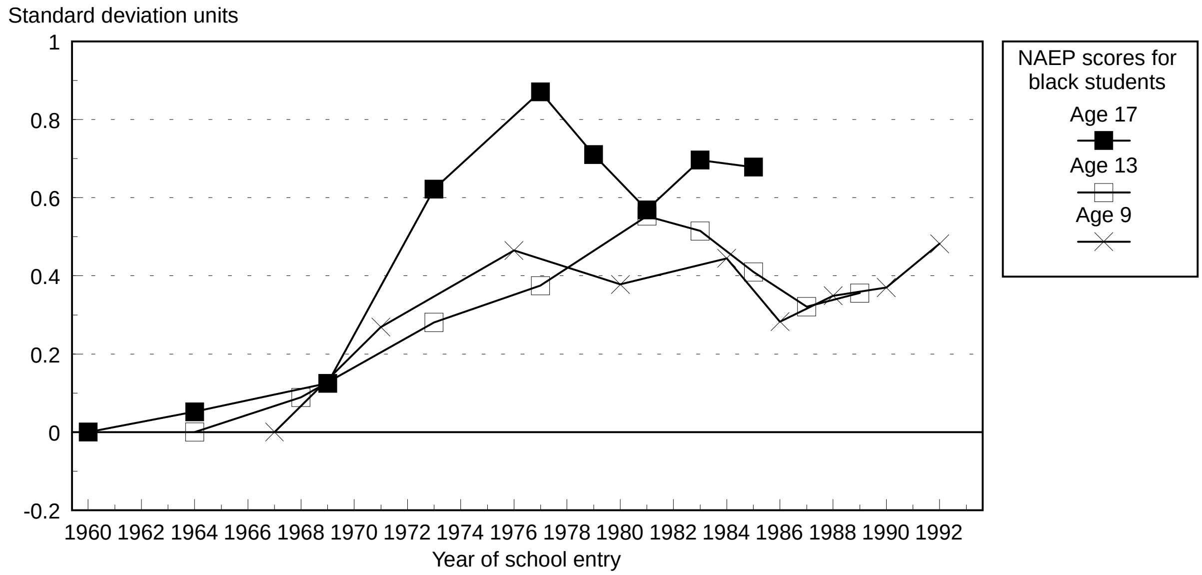
Cohorts. Reading and mathematics are ordinarily thought of as a cumulative process in which early gains are necessary before later gains take place. Charting scores by entering school cohorts tests for the presence this pattern. We characterize cohorts by the year in which they would normally have been in first grade, namely the year in which they were six years old. Each entering school cohort could have taken three tests in their school career—at age 9, 13, and 17. Figures 4 and 5 show each NAEP test score between 1971 and 1996 by the year of school entry. The scores for each age group are connected so that the pattern of increase by age within each entering school cohort can be more easily determined.

The following findings stand out:

- Black gains were small for cohorts entering school prior to 1968.
- The most significant black score gains occurred for cohorts entering school from 1968 through 1972 and 1976 through 1980.[6] After this period of rapid increase in both mathematics and reading scores, mathematics scores have stabilized while reading scores have declined.
- Except for 9-year-old mathematics scores, cohorts entering school after 1980 have registered no further score gains.
- Black reading gains precede mathematics gains. The data show small gains in reading for cohorts entering in the 1960s and dramatic gains for the 1968–72 cohorts. The mathematics data show no evidence of gains before the 1971 cohort.

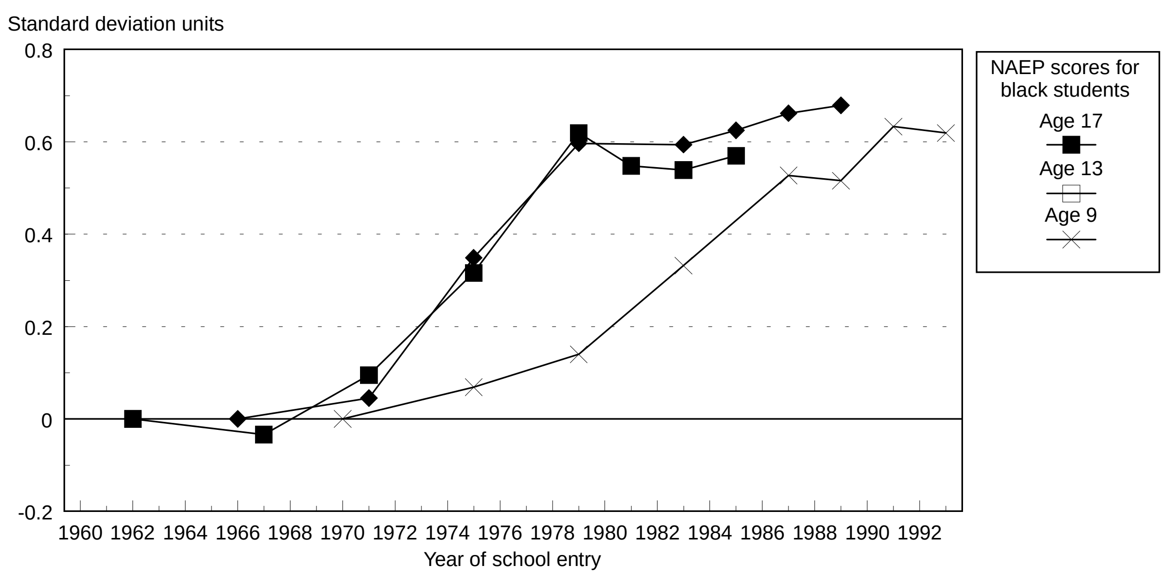
[6] For cohorts entering before 1968, we have no data on 9-year-olds. The position of the cohort curve is thus more uncertain for 9-year-olds than for the 13- or 17-year-olds—especially for reading. The large gain in 9-year-olds' reading between the cohorts entering in 1968 and 1972 may well indicate that there was also some gain at age nine between the 1964 and 1968 cohorts. If that were the case, and if we had the data, it would have the effect of raising all subsequent 9-year-old points in the cohort graphs and make the 9-year-old patterns closer to the pattern for older groups. For mathematics, however, 9-year-olds in the cohorts that entered in 1970 and 1975 scored at about the same level, making earlier gains appear less likely.

Figure 4.—NAEP reading scores for black students, by year of school entry



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 5.—NAEP mathematics scores for black students, by year of school entry



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

While the data does show strong cohort patterns, it also indicates that scores can increase at later ages above gains achieved at earlier ages.

Regions. The regional data shows that significant black gains and black-white gap reductions occurred in all regions for each age and subject, although some regional differences do exist (See figure 6 and 7).[7] Black score gains were somewhat larger in the south and west, although the reduction in the black-white gap was fairly uniform across regions.

Taken together, the NAEP data raise a number of questions:

- Why did both black and white scores rise for all ages in both reading and mathematics?
- Why did black scores rise substantially more than white scores at all ages and in all subjects?
- Why were black gains mainly concentrated for cohorts entering school between 1968–72 and 1976–80?
- Why did older black students gain and then lose more than younger black students?
- Why did black reading gains precede black mathematics gains?
- Why did significant black gains occur in all regions of the country with somewhat higher gains in the south and west?
- Why were black-white gap reductions fairly uniform across regions?
- Why did low-scoring students gain more in mathematics and less in reading than higher scoring students, regardless of race?

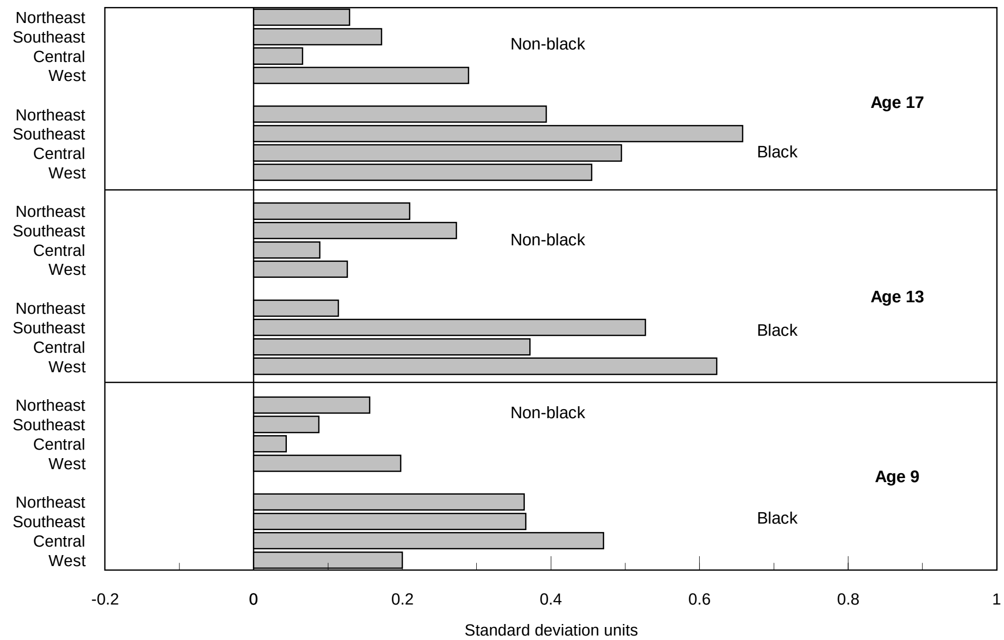
Early childhood interventions are widely thought to have the largest potential effect on academic achievement, partly because of their influence on brain development.

The most striking feature of the NAEP results for blacks is the size of adolescents' gains for cohorts entering from 1968–72 to 1976–1980. These gains were 0.6 standard deviation across subjects. Such large gains for very large national populations over such short time periods in tests similar to the NAEP are rare, if not unprecedented. Scores on IQ tests given to national populations seem to have increased gradually and persistently throughout the twentieth century, both in the United States and elsewhere (Flynn, 1987; Neisser, in press). While evidence exists for large gains on the RAVENS test which measures a narrower ability than tests like the NAEP, the gains on tests similar to the NAEP have averaged about 0.02 standard deviations per year—a fraction of the black rate in the 1980s. Neither these gradual, persistent gains in IQ scores cannot be explained, nor can it be explained whether the gains are larger for minority or other subgroupings of the population (Flynn, 1987). But no evidence exists in this data involving large populations showing gains of the magnitude made by black students over a 10-year period.

It is even unusual to obtain gains of this magnitude in intensive programs explicitly aimed at raising test scores. Early childhood interventions are widely thought to have the largest potential effect on academic achievement, partly because of their influence on brain development. Yet only a handful of "model" programs have reported gains as large as half a standard deviation (Barnett, 1995). These were very small-scale programs with intensive levels of intervention. Even when early childhood programs produce large initial gains, the effects usually diminish over time. Among blacks who entered school between 1968 and 1978 gains were very large among older students and were not confined to small samples, but occurred nationwide.

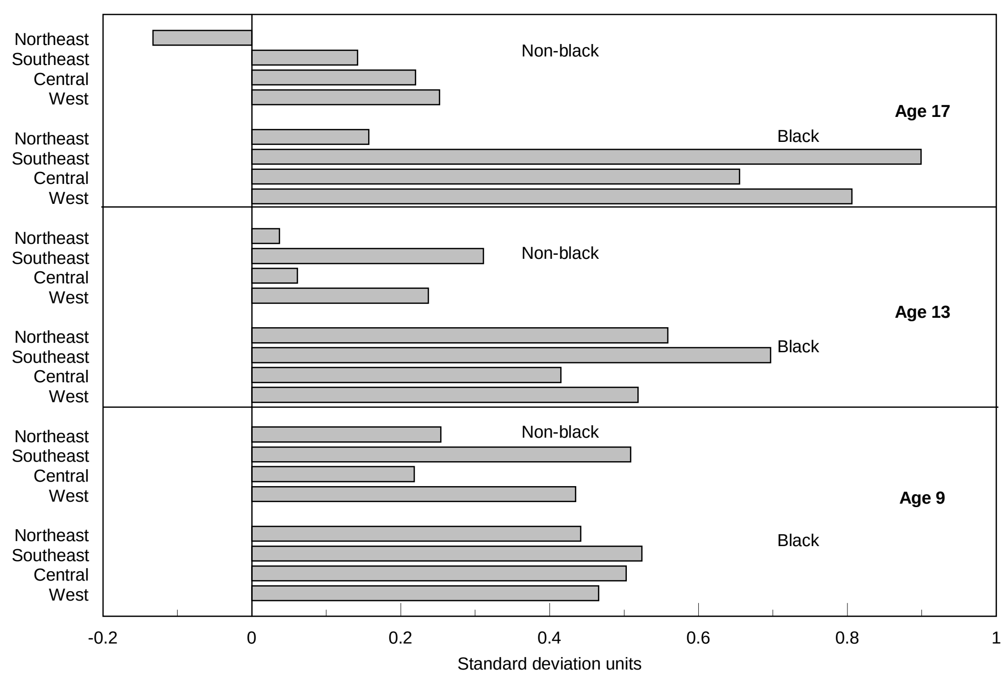
[7] The race x region data is unpublished and was provided by Michael Ross of the National Center for Education Statistics. We only have regional data for 1971–92. For the purpose of reporting NAEP data, the Department of Education places Texas in the West whereas, the U.S. Bureau of the Census places Texas in the South. This is important when interpreting black regional scores in the West, because a sizable proportion of these blacks are in Texas.

Figure 6.—Change in NAEP reading scores between 1971 and 1992, by region, race/ ethnicity, and age



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 7.—Change in NAEP mathematics scores between 1971 and 1992, by region, race/ ethnicity, and age



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Large changes in scores of 0.5 standard deviation and more which are sustained through older ages have been observed when sustained interruptions in schooling occurs at younger ages (Ceci and Williams, 1997). Black students typically gain about 0.4 standard deviations per year on the NAEP tests between the ages of 9 and 13. In terms of "grade equivalents," black adolescent gains were equivalent to approximately 1.5 years of additional schooling. The large black gains sustained for older students suggests that there may have been a major change in the quality of blacks' school experience beginning in the late 1960s. This change in school experiences could reflect social and legal changes aimed at equalizing educational opportunity, additional educational resources that were especially helpful for black students, and the implementation of civil rights legislation creating new job opportunities for academically successful blacks, which may have made black students more eager to take advantage of any opportunities their schools provided.

However, before testing more specific hypothesis about changes in schools, we need to account for how changes in families may have affected test scores. Family characteristics account for the largest part of the variance in test scores in cross sectional models, and family characteristics changed significantly in this period. Thus, it is important to estimate how changing families would be expected to change achievement scores.

Family characteristics account for the largest part of the variance in test scores in cross sectional models . . .

Family Changes

The available evidence would indicate that changes in the family would be expected to have a positive effect on test scores from 1970 to 1990 (Grissmer, et al., 1994) (Cook and Evans, 1998). Higher parental education and smaller family size are the main factors leading to higher predicted test scores of approximately 0.2 standard deviation for black and white students. The sizes of the predicted effects are about the size of the white score gains, but much smaller than the score gains of blacks.[8] While these family gains can account for nearly all white score gains, they can explain only approximately one-third of black gains during the NAEP years.

Therefore, we must turn to events in the educational system: the growth of preschools and kindergartens, desegregation in the South, declines in class size, increases in teachers' age and experience and increases in the amount of teachers' education. Some of these factors would be expected to affect scores only at certain ages or for certain subjects or primarily in certain regions of the country, while others could potentially affect scores nationally at all ages in both subjects.

Changes in Schooling and Educational Resources

An assessment of the impacts of these factors on NAEP scores discounts many of them as substantial contributors to the overall black gains for all age groups (Grissmer et al., 1998).[9] In examining each

8 Even these modest estimates of the gains attributable to improvement in family background may be too high. Consider the case of parental education. Parental education is correlated with children's test performance for two reasons. First, education changes parents in ways that make them more likely to provide their children with an environment conducive to learning. Second, education is a proxy for innate characteristics of parents that they pass along to their children. These innate characteristics also enhance children's test scores. When parents stay in school longer, their child-rearing practices probably change, but their innate characteristics do not. Keeping parents in school longer is, therefore, unlikely to raise children's test scores as much as we would expect on the basis of the cross-sectional estimates.

9 This section presents a summary of much more detailed evidence provided in Grissmer, et al. (1998) for the size of expected effects from the changes in schooling and education cited in the rest of the article.

factor the evidence was assessed for how much a factor changed during the NAEP period, how many youth experienced the change and whether it changed more for blacks, how large the expected effects might be and whether they might be larger for black students, and how well the changes match the changes in NAEP scores.

Kindergarten attendance also increased during this period because of state mandates. A proxy measure of increasing attendance is the percentage of 5-year-olds in school (either kindergarten or first grade). About 66 percent of the entering class in 1960 were in K–1 at age 5 versus 89 percent for the entering class of 1990.[10] There is also a shift toward full- rather than half-day attendance. In 1970 only 12 percent of 5-year-olds attended for a full day, versus 41 percent in 1991.[11] Finally, black participation has increased somewhat faster than white participation. In 1969, 78 percent of white and 67 percent of black 5-year-olds were in K–1 while the percentages were almost equal in 1990.[12] Once again, strong differential effects by race would be required to affect the black-white gap.

The empirical evidence suggests that the large growth in pre-school participation and kindergarten may have a limited impact on 9-year-old scores, but would not significantly impact scores at ages 13 and 17 . . .

The empirical evidence suggests that the large growth in pre-school participation and kindergarten may have a limited impact on 9-year-old scores, but would not significantly impact scores at ages 13 and 17 (Barnett, 1995) (Karweit, 1989). For pre-school the evidence is much stronger due to several methodologically strong studies. The evidence shows that small-scale, intensive interventions can have effects of 0.5 standard deviation or more in the short term, but its effect lessons for almost all studies measured at age 9 or older. Large-scale public programs show even weaker effects with similar fade-out.

The kindergarten evidence is based on differences between half-day and full-day attendance which show significant short-term effects, but similar fade-out effects. Larger short-term effects might be expected from the change from no attendance to half-day attendance, but it would counter the available evidence from both pre-school and full day kindergarten attendance for long-term effects to result from such a change. Similar to pre-school, more kindergarten participation might have residual effects to age 9, but no longer-term effects would be expected.[13]

The percentage of black high school graduates completing a minimum set of specified courses (4 years of English, 3 years of social science, 2 years of science and 2 years of mathematics) increased from 32 percent in 1982 to 76 percent in 1994. However, gains for white students were similar—from 33 to 76 percent. Changes in course work at the high school level in the 1980s may explain part of the score increases for older black and white students, but probably cannot explain much of the differential black score gains since course work changes were similar for black and white students.

Desegregation also appears to offer an explanation for a small part of the black score gains for all ages. Desegregation occurred primarily in the south over a short period in the late 1960s and early 1970s, but the regional NAEP data shows that black score gains occurred in all regions of the country. While the largest gains appear to have been made in the south, the extra southern gains accounts for less than 20 percent of overall black gains.

---

[10] Ideally, a measure of the percentage of those entering first grade who attended kindergarten is needed. By 1990, over 98 percent of children entering school attended kindergarten. However, such data is not available for earlier years. Our measure does not approach 98 percent because an increasing number of children in the 1980s delayed school entry for one year and attended kindergarten at age 6.

[11] Data from *Digest of Education Statistics*, 1992, table 47.

[12] *Digest of Education Statistics*, 1970, table 4 and *Digest of Education Statistics*, 1992, table 7.

[13] Similar to preschool also, some small-scale, specially designed kindergarten programs appear to have substantial short-term effects (Karweit, 1989). It is possible that kindergarten curriculum has shifted with some effects at age 9 as well.

Nationwide, three significant changes took place in schools during this period—lower class size, and better educated and more experienced teachers. Unlike many of the changes cited above, these changes, would have been experienced by nearly all students at all ages. Thus, if such changes would be expected to have effects on achievement, these changes would better explain NAEP score gains for all age groups and subjects.

Teachers' average level of experience declined in the early 1960s as substantial numbers of inexperienced teachers were hired to teach the baby boomers. As enrollments fell in the 1970s, the flow of new, inexperienced teachers was substantially reduced; the average experience level of teachers grew substantially from 1970 to 1990 (Grissmer and Kirby, 1997), and, by 1990, a significant number of teachers had 20 or more years of experience. We assume here that teachers as a group are most productive between 5 and 20 years of experience. Figure 8 shows the average changes in the percentage of teachers with 5–20 years of experience who would be teaching an entering cohort of age 9, 13, or 17 children.[14] The percentage of teachers in this experience range grew considerably for cohorts entering from the 1960s to the mid-1980s, then fell with the growth of teachers with 20 or more years of experience.

Nationwide, three significant changes took place in schools during this period—lower class size, and better educated and more experienced teachers.

Along with this gain in experience came more teachers with master's degrees. Figure 9 shows that the percentage of teachers with master's degrees experienced by an entering school cohort rose for all age groups. The education level of teachers grew steadily for entering school cohorts from the 1960s to the 1990s—although the period of fastest growth was from the 1960s to the mid-1980s. The size of classes was also reduced substantially during this period. Figure 10 shows the average pupil/teacher ratio for entering school cohorts up to age 9, 13, and 17. The pupil/teacher ratio—a measure of class size—also fell dramatically for cohorts entering school in the 1960s and 1970s, but slowed considerably in the 1980s. These changes occurred at both elementary and secondary levels although the timing was somewhat different. Part of the reason class sizes fell was also related to the baby boom. As enrollments dropped in the 1970s, rather than terminate teachers, the opportunity was used to reduce class size.

The empirical evidence on the effects of these three variables more greatly impacts class size than teacher experience or education for two reasons. First experimental evidence exists for the effects of class size. Second, the accuracy of all measurements using non-experimental data is now being questioned.

A large, multi-district study in Tennessee that randomly assigned students to classes of approximately 14 students instead of approximately 22, found that reducing class size between kindergarten and third grade had significant effects on achievement, and even greater effects for blacks (Krueger, 1997; Mosteller, 1994). The effects averaged about 0.20 standard deviations for whites and 0.30 standard deviations for blacks, with equal effects in reading and mathematics. Following the experiment, Tennessee also cut class sizes in 13 school districts with the lowest family income. Comparisons with other districts and test score changes within these districts showed gains of 0.35 to 0.5 standard deviations (see Mosteller, 1994).[15] The Tennessee data suggest that disadvantaged students may experience the most gains from class size reductions.

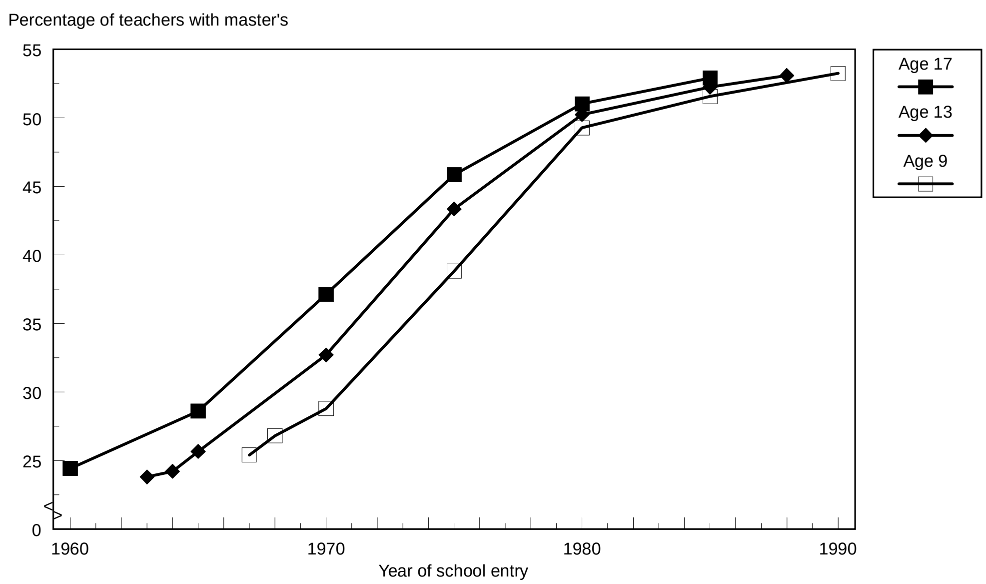
[14] We somewhat arbitrarily chose 5–20 years of experience as the period of peak productivity for teachers assuming an early learning curve for teachers and some average "burnout" effect after 20 years of service. The data here is the average percentage of teachers with 5–20 years of experience during the schooling experience of each age group from school entry. Thus 9-year-old students who entered school in 1970 would be estimated by taking the average percentage of teachers with 5–20 years of service between 1970 and 1973. For 13-year-old students entering in 1970, the average would be from 1970 to 1977.

[15] Other research on school districts in Alabama shows similar overall effects when using models without prior year test score controls (Ferguson and Ladd, 1996).

Figure 8.—Average percentage of teachers with between 5–20 years of teaching experience in years of attendance

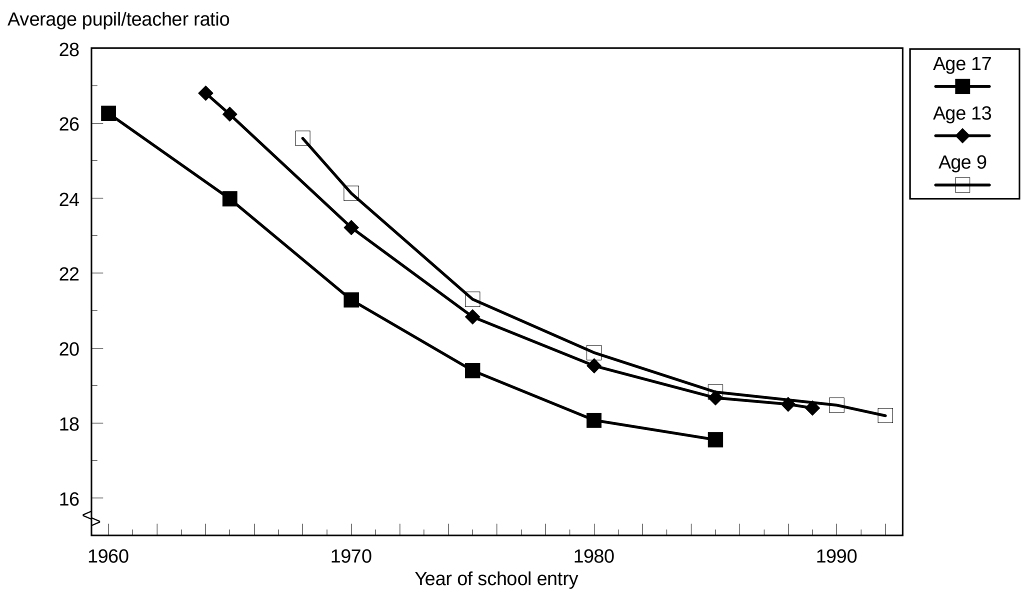
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 9.—Average percentage of teachers with a master's degree or higher for years of attendance



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

Figure 10.—Average cohort pupil/teacher ratio for years of attendance, by year of school entry



SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP).

It is likely that had a similar experiment been done using a national sample of K–3 students in the 1970s, even larger differences between black and white test scores would have been measured. This is because white students in Tennessee are poorer than average white students nationwide and black students in Tennessee probably were less disadvantaged in 1990 than blacks nationwide in the 1960s and 1970s. National reductions in class size in the 1960s and 1970s were approximately the same as the Tennessee experiment, thus effects ranging from 0.3 to 0.6 might be expected for black students and 0.00 to 0.20 for white students from the national class size reductions. Thus, reductions in class size may be a key factor in explaining large black gains and why black scores may have risen much more than white scores.

The Tennessee experimental evidence also leaves a number of questions unanswered. First, we do not know much about either the long-term effects of smaller elementary school classes or the cumulative effects of smaller classes from kindergarten through twelfth grade. In the Tennessee experiment, students were returned to large classes after third grade. By seventh grade the standardized benefits of smaller classes were only half as large as they had been at the end of third grade, and the benefits to black students were no larger than the benefits to whites (Mosteller, 1995). We do not know what would have happened if classes had remained small until students finished school. Second, the Tennessee measurements may only represent a short-term effect since only a single cohort was measured. Teachers and policymakers may be able to adapt their teaching and policies to take better advantage of smaller class sizes in the longer term.

Another problem with the hypothesis that class size reductions raised test scores is that class size fell in the 1960s as well as the 1970s. If smaller classes had conferred long-term benefits, 17-year-olds who entered school in 1968 should have outscored those who entered in 1960. This was not the case outside the South.[16] Further research is required to test the class size hypothesis as a strong contributing factor to black gains.

One important side effect of the Tennessee experiment is that it raised new doubts about non-experimental studies conducted in a "production function" framework.[17] These studies typically try to control standardized scores at Time One and then discern whether a resource like smaller classes affects gains between Time One and Time Two. In Tennessee, however, smaller classes exerted their entire effect on standardized scores in the first year. Thereafter, smaller classes simply served to sustain the initial standardized gains. Thus, the estimated effect on smaller classes in grades one through three would have been zero.

Current empirical measurements of the effects of teacher education and a master's degree show no consistently strong effects—but better specified models might change these results. It remains to be seen whether more teacher education and more experience raised achievement scores awaits stronger empirical evidence and determination of the current flaws in specifying estimation models.

Discussion

Recent research is undermining several of the assumptions and empirical evidence underlying the "money doesn't matter" conclusion. The validity of the empirical studies reviewed to arrive at this conclusion is being questioned due to the use of model specification which would not reproduce the experimental class size results. It is possible that fundamental flaws are present in nearly all non-experimental studies of the effects of school resources due to the methods of model specification. Second, experimental data which avoids the assumptions needed in models with non-experimental data indicates that reductions in class size—a key school resource parameter—have significant effects with larger effects for minority students. Third, NAEP data—which had previously been used together with the large perceived increases in school resources to support the "money doesn't matter" argument—now seems more supportive of a different conclusion. This evidence seems to support the thesis that money directed at minority and disadvantaged students brings higher achievement scores, but money directed toward more advantaged students may have much smaller or negligible effect. Moreover, the additional money available in the 1960s to 1990s was much less than previous estimates. Instead of doubling in real terms, the real increases directed toward achievement of regular students was closer to 30 percent during this period. These additional resources were also disproportionately directed toward minority and lower income students. Thus, a more consistent story is emerging from the empirical data which is more supportive of the thesis that additional money matters greatly for minority and disadvantaged students, but much less or little for advantaged students.

[16] Class size effects may depend on how teachers change their behavior when they have smaller classes (Murnane, 1996). The effects of smaller classes may take several years to appear, because teachers and students need time to adjust their behavior to smaller classes. However, effects appeared immediately in Tennessee and increased very little as a result of additional years in small classes. The Tennessee experiment did not address the important question of whether effects would grow at each grade level as teachers experienced smaller class sizes over many years.

[17] The Tennessee data shows that gains from smaller classes appear immediately and grew by small amounts over the first three grades (Krueger, 1997). This implies that production functions that utilize previous year's scores would not measure the class size effects evident in Tennessee. Two studies that have used some of the best data at the state level and had prior year's test scores as controls were considered to be among the strongest studies (Ferguson, 1991 and Ferguson and Ladd, 1996). However, these specifications now appear to provide biased results. In the latter study, results are also presented of cross-sectional estimates without controls, and these results may now have more credibility than those with prior year controls. Generally, the results of the model without prior score controls show stronger effects for most variables.

References

Campbell, Jay, Kristin Voelkl, and Patricia Donahue. *NAEP 1996 Trends in Academic Progress*, Washington, D.C.: National Center for Education Statistics / Educational Testing Service, NCES 97–985, September 1997.

Ceci, Stephen, and W. Williams. Schooling, Intelligence, and Income. *American Psychologist*, vol. 52: 1051–1058, 1997.

Cook, Michael, and William N. Evans. "Families or Schools? Explaining the Convergence in White and Black Academic Performance," draft copy.

Ferguson, Ronald F. "Paying for Public Education: New Evidence on How and Why Money Matters." *Harvard Journal on Legislation*, vol. 28: 465–97, 1991.

Ferguson, Ronald F., and Helen F. Ladd. "How and Why Money Matters: An Analysis of Alabama Schools." In Helen F. Ladd, ed. *Holding Schools Accountable.* Washington, D.C.: The Brookings Institution, 1996.

Flynn, J. R. "Massive IQ gains in 14 nations: What IQ tests really measure." *Psychological Bulletin*, vol. 101, no. 2: 171–191, 1987.

Gamoran, Adam. "Student Achievement in Public Magnet, Public Comprehensize, and Private City High Schools." *Educational Evaluation and Policy Analysis*, vol. 18, no. 1: 1–18, Spring 1996.

Grissmer, David, A. Flanagan, and S. Williamson. "Why Did Black Test Scores Rise Rapidly in the 1970s and 1980s." In *The Black-White Test Score Gap,* Christopher Jencks and Merideth Phillips, eds., Washington, D.C.: The Brookings Institution, Forthcoming.

Grissmer, David, and Sheila Nataraj Kirby. "Teacher Turnover and Teacher Quality," *Teacher's College Record*, vol. 99, no. 1: 45–56. Fall 1997.

Grissmer, David , Sheila Nataraj Kirby, Mark Berends, and Stephanie Williamson. "Exploring the Rapid Rise in Black Achievement Scores in the United States (1970–1990)." In *The Rising Curve: Long-Term Changes in IQ and Related Measures,* U. Neisser, ed. Washington, D.C.: American Psychological Association, 1998.

Grissmer, D. W., S. N. Kirby, M. Berends, and S. Williamson, 1994. Student Achievement and the Changing American Family, Santa Monica, CA: RAND.

Hanushek, Eric A. "School Resources and Student Performance." In D*oes Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, G. Burtless, ed. Washington, D.C.: The Brookings Institution, 1996.

Hanushek, E. A. *Making Schools Work: Improving Performance and Controlling Costs.* Washington, D.C.: The Brookings Institution, 1994.

Hanushek, E.A. "The Impact of Differential Expenditures on School Performance." *Educational Researcher*, vol. 18, no. 4: 45–51, 1989.

Hedges, Larry V., and Amy Nowell. "Group Differences in Mental Test Scores: Mean Differences, Variability, and Talent." In C.S. Jencks and M. Phillips, eds. *The Black-White Test Score Gap.* forthcoming 1998.

Hedges, Larry V., and Rob Greenwald. "Have Times Changed? The Relation between School Resources and Student Performance." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, G. Burtless, ed. Washington, D.C.: The Brookings Institution, 1996.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. "Does Monday Matter: Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher*, vol. 23, no. 3: 5–14, 1992.

Karweit, Nancy. "Effective Kindergarten Programs and Practices for Students at Risk." In *Effective Programs for Students At Risk*, Slavin, Robert, Karweit, N. and Nancy Madden. Boston: Allyn and Bacon, 1989.

Krueger, Alan B., Experimental Estimates of Education Productions Functions, Princeton, NJ: Princeton University, Industrial Relations Section, Working Paper #379, May 1997.

Ladd, Helen F. *Holding Schools Accountable.* Washington, D.C.: The Brookings Institution, 1996a.

Ladd, Helen F. "Introduction." In *Holding Schools Accountable*, H.F. Ladd, ed. Washington, D.C.: The Brookings Institution, 1996b.

Lankford, Hamilton and James Wyckoff, "The Allocation of Resources to Special Education and Regular Instruction." In *Holding Schools Accountable*, Helen F. Ladd, ed. Washington, D.C.: The Brookings Institution, 1996.

Mosteller, F. "The Tennessee Study of Class Size in the Early School Grades." *The Future of Children*, vol. 5, no. 2: 113–127, 1995.

Murnane, Richard J., and Frank Levy. "Evidence from Fifteen Schools in Austin, Texas." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, G. Burtless, ed. Washington, D.C.: The Brookings Institution Press, 1996.

Neisser U., ed. *The Rising Curve: Long-Term Changes in IQ and Related Measures.* Washington, D.C.: American Psychological Association, 1998.

Rasinski, Kenneth, and Steven Ingels. "America's High School Sophomores: A Ten Year Comparison." NCES 93-087, Washington, D.C.: National Center for Education Statistics, June 1993.

Raudenbush, Stephen W., Randall P. Fotiu, and Yuk Fai Cheong. "Inequality of Access to Educational Resources: A National Report Card for Eighth Grade Math," forthcoming.

Rothstein, R., and K.H. Miles. *Where's the money gone? Changes in the level and composition of education spending.* Washington, D.C.: Economic Policy Institute, 1995.