

Comprehensive School Reform: Meta-Analytic Evidence of Black-White Achievement Gap Narrowing¹

Kevin M. Gorey

University of Windsor, Canada

Abstract

This meta-analysis extends a previous review of the achievement effects of comprehensive school reform (CSR) programs (Borman, Hewes, Overman, & Brown, 2003). That meta-analysis observed significant effects of well endowed and well-researched programs, but it did not account for race/ethnicity. This article synthesizes 34 cohort or quasi-experimental outcomes of studies that incorporated the policy-critical characteristic of race/ethnicity. Findings: compared with matched traditional schools, the black-white achievement gap narrowed significantly more among students in CSR schools. In addition, the aggregate effects were large, substantially to completely eliminating the achievement gap between African American and non-Hispanic white students in elementary and middle schools. Title I policies before or after the No Child Left Behind Act of 2001 seem to have had essentially no impact on the black-white achievement gap. Curricular and testing mandates along with the threat of sanctions without concomitant resource supports seem to have failed. This study suggests that educational achievement inequities need not be America's destiny. It seems that they could be eliminated through concerted political will and ample resource commitments to evidence-based educational programs.

Keywords

comprehensive school reform; Title I; racial differences; academic achievement; education policy; elementary schools; middle schools; meta-analysis

Introduction

Federal funding support for schools serving relatively low-income areas to bolster the educational opportunities of children at risk of under-achieving has been available for more than 40 years. Title I of the Elementary and Secondary Education Act of 1965 initially supported a variety of supplemental services that were typically internally developed by school districts and enacted as diverse so-called pull-out programs. In the mid-1990s, Title I

¹This research was supported in part by a Canadian Institutes of Health Research investigator award and a grant from the Social Sciences and Humanities Research Council of Canada.

Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives (Archivos Analíticos de Políticas Educativas)**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA/AAPE**. **EPAA/AAPE** iqs published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by the Directory of Open Access Journals, H.W. Wilson & Co., and SCOPUS.

reauthorization encouraged school-wide initiatives at the same time that the Comprehensive School Reform Demonstration Program (CSRP) supported the institution of externally developed, scientifically-based, school-wide comprehensive programs. Dozens of such Comprehensive School Reform (CSR) models have since been developed. Recent Title I reauthorization—the No Child Left Behind Act (NCLB, 2002)—brought Title I and CSRP under the same legislative umbrella. Historically, Title I had always targeted educationally or otherwise deprived children, who regrettably remain more prevalently represented among people of color in America, but NCLB explicitly set as one of its central goals the elimination of academic achievement gaps by race/ethnicity. Contemporary underfunding criticisms notwithstanding, Title I programs, represented in tens of thousands of schools, have enjoyed hundreds of billions of dollars in federal support thus far. One may fairly wonder how effective any such school interventions, piecemeal or more comprehensive, have been in narrowing the well documented black-white academic achievement gap. A confident answer to that rhetorical question does not yet seem to exist. Previous meta-analytic reviews of the achievement effects of internally developed Title I programs and externally developed CSR programs (Borman & D’Agostino, 1996; Borman, Hewes, Overman, & Brown, 2003), respectively, found very modest to strong overall effects; however, they did not account for race/ethnicity. This meta-analysis will do so by focusing specifically on the effects of such educational programs on the black-white achievement test score gap.

The Black-White Test Score Gap

Because academic achievement can open the door to so many of society’s opportunities, the black-white achievement gap may be thought of as a sentinel indicator of society’s remnant injustice. Its disappearance one day will mark the end of the nation’s civil rights march, our arrival at the “promised land” that Dr. Martin Luther King, Jr., dreamed of. A very large gap in the academic achievement test performance of black and white children existed in America’s elementary, junior or middle, and senior high schools in the 1960s. It typically ranged from a half a standard deviation (*SD*) deficit among black children in elementary school to more than a full *SD* difference by 12th grade. Gratefully, in seeming correlation with the nation’s war on poverty and development of such compensatory early educational programs as Head Start, the gap narrowed by about one tenth of a *SD* per decade for thirty years or so (Grissmer, Flanagan, & Williamson, 1998; Hedges & Nowell, 1998). Regrettably, that trend began to slow and even to reverse in some places through the 1990s, and previous gains seem essentially to have flat-lined during the NCLB era (Lee, 2006). Controversial genetic explanations for such racial group differences are well known (Herrnstein & Murray, 1994; Rushton & Jensen, 2005), but they seem shallow in the American social context, even farfetched (Gorey, 2001; Gorey & Cryns, 1995; Gorey et al., 2009). For example, any such population-level, between-race genetic differences have not changed systematically over the mere two generations since the Great Society’s reforms. However, social policies affecting households, neighborhoods, and schools have changed quite a bit and seemingly contemporaneously with changes in the black-white test score gap. Such social policy-social effect parsimony provides great hope—causes of the black-white test score gap seem malleable and amenable to effective intervention.

Parental, household, and neighborhood factors can probably account for between 25% and 50% of the black-white achievement gap in the United States. Key interrelated predictors seem to be neighborhood segregation and other neighborhood factors associated with poor health, low parental education and income status, and their household correlates that indicate few learning supports and stimulations (Brooks-Gunn, Klebanov, Smith, Duncan, & Lee, 2003; Card & Rothstein, 2007; Hedges & Nowell, 1998; Magnuson & Duncan, 2006; Rothstein, 2004). Clearly, schools could not have singularly caused nor can they ever hope to be a singular solution to the black-white achievement gap. However, in that they are places mandated to support academic development where children typically spend 30 or more hours a week, schools are probably an important component cause of the gap. The achievement test scores of African American and white children diverge by one tenth of a *SD* per year throughout elementary school, and this divergence seems to be related to the relatively poor quality of prevalently segregated schools (Fryer & Levitt, 2004, 2006). Traditional school programs have demonstrated the ability to significantly diminish cognitive gaps associated with socioeconomic status, gender and even other “unexplained” gaps, but not those associated with race (Downey, von Hippel, & Broh, 2004; Entwisle & Alexander, 1994). It seems plausible that more comprehensive programs that offer increased instructional quality would do better. Questions about the strength of school quality-race achievement gap associations remain largely unanswered. They could be quite large. This review aims to begin to answer such questions related to two instructional program enhancements: Title I and CSR.

Title I, Comprehensive School Reform and Achievement

Meta-analyses of hundreds of elementary and middle school study outcomes have demonstrated unequivocally that educational resources matter, particularly if they are well placed, producing smaller classes taught by better trained and more experienced teachers. Relatively modest per-capita funding increases (e.g., \$500 per child) have been associated with significant overall achievement gains (up to a one-quarter *SD* gain on standardized achievement tests [Glass & Smith, 1979; Greenwald, Hedges, & Laine, 1996; Hedges, Laine, & Greenwald, 1994]). Recently, two studies with secondary analysis of an aggregated 10 state databases from 1985 to 2000 have suggested that funding policies that positively affect class size and teacher quality may also narrow the black-white achievement test score gap (Braun, Wang, Jenkins, & Weinbaum, 2006; Krueger & Whitmore, 2001). Resources that affect human capital in education clearly seem to make a difference. What of resources that affect instruction? They also probably matter very much. For example, one review of one specific comprehensive high school reform found a significantly increased graduation rate among African American men and conservatively estimated that its public benefits (through increased tax revenues and decreased costs related to illness and crime) would be nearly five times its costs (Levin, Belfield, Muennig, & Rouse, 2007; Quint, Bloom, Rebeck Black, Stephens, & Akey, 2005).

Previous Meta-Analyses on Instructional Resources

Two exhaustive meta-analytic reviews of the achievement effects of internally developed, often relatively piecemeal pull-out Title I programs and externally developed, more comprehensive school-wide CSR programs, respectively, found modest to moderate overall

effects (nearly 2,000 outcomes of elementary and secondary instructional intervention programs, 1965 to 2001 [Borman & D'Agostino, 1996; Borman et al., 2003]). On average, 60% of children in CSR schools scored better on standardized academic achievement tests than their typical counterpart in a traditional school, while only 55% of Title I supported children scored above their typical non-supported counterpart. This difference is perhaps not surprising as CSR programs enjoy much greater funding support. Instructional resources also mattered significantly in these meta-analyses. On average, longer-standing and better endowed programs demonstrated larger effects. In fact, the average achievement effect of CSR programs that had been implemented for five to seven or more years could fairly be characterized as quite strong. Nearly three-quarters of their students scored better than a typical student in a traditional school. Even though these previous meta-analyses were not able to account for race/ethnicity, because the vast majority of their aggregate samples were minority children of color or eligible for school lunch programs, they did provide indirect suggestive evidence, particularly for the more effective CSR programs, that they probably also serve to narrow the black-white test score gap. This meta-analysis aims to extend the synthetic knowledge produced by previous meta-analyses by developing and examining that notion. In light of this field's historical, theoretical and empirical contexts, it explores whether Title I programs significantly narrow the black-white achievement gap and whether CSR programs narrow the gap even more

Methods

Selection of the Sample for Meta-Analysis

Initial informal scans of this field's published research literature found few empirically relevant studies. It seemed that most of the conceptually relevant studies did not report their findings in enough detail to allow for the calculation of effect sizes corresponding to black-white test score gaps separately for pre- and post-Title I cohort sub-samples or for CSR and traditional school sub-samples. Typically, that analysis requires authors to report separate African American and non-Hispanic white achievement test mean scores and separate sub-sample *SDs* or their statistical analogues (e.g., *t*-tests or *F*-ratios). As such analytic reporting detail seemed more characteristic of dissertation research, dissertations were chosen as this meta-analysis's central sampling frame.

In June of 2009 the *Dissertation Abstracts International* database was searched from 1985 to 2009 using the following key word search scheme: (comprehensive school reform [CSR] or names of the 12 most well-researched program models [Direct Instruction, School Development Program, Success for All, Expeditionary Learning Outward Bound, Modern Red Schoolhouse, Roots and Wings, Accelerated Schools, America's Choice, ATLAS Communities, Montessori, Paideia or The Learning Network; Borman et al., 2003] or Elementary and Secondary Education Act [ESEA], Title I or No Child Left Behind [NCLB]) and (race, ethnic*, black* or African-American). This search was augmented with searches of the following published and unpublished research literature databases: *Education Resources Information Center (ERIC)*, *PsycINFO* and the *Conference Papers Index*, then with bibliographic reviews of retrieved manuscripts and World Wide Web searches of CSR developers' and relevant state and federal education agency websites. Additional inclusion

criteria: an assessment of school-wide rather than pull-out instructional interventions; a provision of research methodological control for major potential confound explanations such as history, maturation and regression toward the mean; analysis and reporting of achievement test score data. Discrete outcome assessments (e.g., pass/fail) were excluded because their diverse criterion cut-off definitions make them incomparable between studies.

Twelve studies met all of these conceptual, methodological and empirical criteria: eight dissertations, two journal articles, one government document, and one non-profit agency report (marked with an asterisk in the references section). These 12 selected studies tested 17 paired hypotheses (e.g., testing narrowing of the black-white test score gap in traditional schools and in CSR schools), providing 34 study outcomes for this meta-analysis. If a study used more than one measure of the same conceptual achievement domain, the effects were combined (with a weighted average) into one independent hypothesis test. If a study used multiple conceptual measures (e.g., language arts and mathematics), the measures were treated as independent hypotheses for this meta-analysis.

Sample Description

This review of the practical black-white achievement gap impacts of CSR in the United States is based on 34 study outcomes of instructional programs (17 outcome pairs) initiated in 415 elementary or middle schools between 1985 and 2005. Studies typically produced two outcomes (e.g., black-white achievement gap in traditional schools and in CSR schools). Most of the research was accomplished in urban schools whose aggregate participant sample of 125,465 included 22,001 African American children. Half of the study samples were based in a single city school district, while the other half, state-wide or national samples, were also largely representative of urban areas. Two suburban and two rural samples were also included. Free-lunch eligibility, a proxy for family low-income status, ranged from 35% to 90%, but typically three-quarters of the children in sampled schools were so eligible. Additional information is presented in Table 1.

As for the school interventions, nine paired study outcomes were of Title I programs and eight paired outcomes were of CSR programs (five Success for All, and one each Direct Instruction, Montessori, and Paideia). Two of the studies of Success for All were conducted by its program developers. All of the other studies were authored by more disinterested external investigators. Studies were fairly large (a median of 700 participants), rigorous, and intensive. The typical study of Title I programs used a controlled retrospective cohort research design while CSR programs were typically quasi-experimentally assessed through comparisons with matched traditional schools. The duration of CSR programs ranged from one to four years (median, 3.0) and they were followed from one to six years (median, 3.0). All of the studies used standardized measures of reading, language arts, or mathematics. Because the observed black-white achievement effects did not differ significantly between these three conceptual definitions of academic achievement, they were aggregated in this meta-analysis.

Meta-Analysis and Effect Size Interpretation

The Cohen (1988) d -index served as the meta-analysis central effect size statistic. Allowing for the translation of the primary studies' diverse statistical outputs into a common metric, it thus aids in making between-study comparisons. It can be calculated directly from study group means and SD s,

$$d = \frac{M_1 - M_2}{(SD_1 - SD_2)/2},$$

or derived from a host of parametric statistics (Cooper, 1998). Related to this study's key concepts, it characterizes the size of the reported black-white achievement test score gap. For example, a d of 0.50 in the hypothesized direction would mean that, on average, the African American children in the primary study in question scored one half of one standard deviation (SD) lower than their non-Hispanic white counterparts on the standardized academic achievement measure used in the study. Cohen's (1988) U_3 statistic augmented d -indexes. An intuitively appealing metric, it compares all of the scores of one study group's with the average score of another's. Again, it aids across and between-study comparisons, but in addition, as it tends to put the emphasis on people rather than on measures or mere aggregate statistics, it seems to lend itself well to the assessment of practical policy significance. For example, a U_3 of 66% resulting from a 3-year post-test comparison of African American and non-Hispanic white children in a traditional elementary school on a standardized measure of academic achievement would be interpreted as follows: two-thirds of the African American children scored lower on the achievement test than the typical white child did.

In pooling effects, this meta-analysis used fixed effects models, as it assumed homogeneity of variance within specific categories of interest (Cooper, 1998): Title I programs at initiation and follow-up and CSR and traditional school programs at follow-up. The black-white test score gap was significantly smaller among middle school samples (middle school programs were significantly more mature), so elementary and middle schools were assessed separately. Within each key categorical distribution, effects (d s) were weighted by their inverse variances so that larger, more precise study outcomes influenced the summary measures more than smaller, less precise study outcomes. And 95% confidence intervals (CI) were constructed around each average effect size estimate (Greenland, 1987; Hedges & Olkin, 1985). Each effect size (d -index) distribution was then tested for heterogeneity using Cochran's Q statistic (Fleiss, Levin, & Paik, 2003). It has a chi-square (χ^2) distribution, which allows one to test if the variability of study effects is greater than would be expected by random variability alone. If so, possible sources of contextual, programmatic, and research methodological variability were explored with meta-regression models. Finally, differences between meta-analytic group means (e.g., comparison of the black-white test score gap in CSR and traditional schools) were tested with an extension of Cochran's Q , the Q_b statistic, which also has a χ^2 distribution. It is essentially the meta-analytic analogue of the t -test or F -ratio.

Results

Results are summarized in Table 2. The black-white achievement gap did not seem to narrow at all with Title I program interventions (top half of table). In elementary schools the one-half standard deviation (*SD*) gap was essentially unchanged from the beginning to the end of cohort follow-up periods; $\chi^2(7) = 2.21$, not significant (*NS*). In middle schools a somewhat smaller gap of between one quarter and one third *SD* was similarly maintained; $\chi^2(9) = 9.82$, *NS*. Better resourced, comprehensive programs appear to have been much more effective in narrowing the black-white achievement gap (bottom half of table). The black-white achievement test score gap decreased significantly more in CSR elementary ($d = 0.15$ [95% CI 0.12, 0.18]) and middle schools (0.06 [−0.01, 0.13]) than in matched traditional schools (0.64 [0.61, 0.67] and 0.21 [0.14, 0.28]); respectively, $\chi^2(11) = 657.23$, $p < .001$, and $\chi^2(3) = 8.91$, $p < .05$. Critical post-intervention comparisons found the typical one-half *SD* black-white achievement gap to have decreased not only statistically but practically in the CSR elementary schools studied to only slightly more than one tenth of a standard deviation.

In terms of the raw numbers of children affected, in traditional elementary schools an estimated 15 of every 20 African American children scored below the typical non-Hispanic white child on standardized measures of academic achievement ($U_3 = 73.9\%$). That estimate was only 11 of every 20 in CSR schools ($U_3 = 55.9\%$). It even seems, though a very tentative inference based on two Success for All study outcomes, that by middle school the achievement test performance gap of African American and white children is far lower in CSR schools ($d = 0.06$ [−0.01, 0.13], *NS* as the CI includes the null value of 0.00). Finally, of the eight effect-size distributions displayed in table 1, only the one associated with traditional elementary schools demonstrated significantly more heterogeneity than expected; $\chi^2(5) = 32.94$, $p < .001$. And none of the contextual (time and place), school (prevalence of African American children and free lunch eligibility), program (grades participating and duration) or research design (sample size, type comparison condition and length of follow-up) characteristics coded in this meta-analysis were significantly associated with its effect variability.

Discussion

This exploratory meta-analysis, primarily of dissertation research, found that externally developed, science or evidence-based comprehensive, school-wide instructional interventions (CSR) can significantly narrow the black-white achievement gap among elementary and middle school students. No such significant achievement gap narrowing was found among less comprehensive, Title I program variations that were typically locally developed. These central findings were consistent with the moderate to strong overall achievement effects of CSR and the very modest overall effects of Title I programs found by previous meta-analysts (Borman et al., 2003; Borman & D'Agostino, 1996). This integrative review's positive CSR findings, most representative of one CSR model—Success for All—were also consistent with two previous narrative reviews of that model's affect on the black-white and Hispanic-non-Hispanic achievement gaps (Slavin & Madden, 2001, 2006). The previous meta-analyses had also observed that more mature Title I programs and more long-

standing CSR interventions (5–7 or more years) were more effective in raising achievement levels. This study lacked sufficient meta-analytic power to systematically replicate such resource-outcome relationships as its relatively small sample of Title I and CSR programs were quite homogeneous on these scores. All of its sampled Title I programs were initiated between 1997 and 2005 (only one before 2000), and all of its sampled CSR interventions had been in place between two and four years. These limits suggests that this study's already positive estimates of CSR effects could underestimate results as none of this field's longest-standing CSR interventions were represented. This review's simple dose-response relation (CSR interventions are significantly better endowed and more effective) still seems to allow for the inference that instructional funding support matters. Mere Title I policies themselves seem to have essentially no impact on the black-white achievement gap. Curricular and testing mandates along with the threat of sanctions without concomitant instructional resource supports, hallmarks of the NCLB era, seem doomed to failure. On the other hand, well-funded, rationally-planned, comprehensive instructional programs seem likely to succeed.

Meta-Analytic and Primary Research Limitations

Some may be concerned with the dissertation sampling frame of this meta-analysis. If any consequent bias did intrude, it probably bodes for the key finding of this meta-analysis being an underestimate. That is, the observed black-white achievement gap decrease due to CSR intervention may be even larger. It has been consistently found, not only in educational research but across other professional and social science disciplines, that the hypothesized effects observed in various unpublished forums, including dissertations, are smaller on average than published effects (de Smidt, 1997; Grenier & Gorey, 1998; Staines & Cleland, 2007; Torgerson, 2006). In fact, this review's focus on dissertation research would seem to provide substantial protection against any threat of publication bias. It should also be noted that this review's meta-analytic question concerning narrowing of the black-white achievement gap was not the primary hypothetical concern among the majority of its included studies. It was most typically a secondary or exploratory hypothesis, and a few studies merely reported relevant data that was amenable to this meta-analysis' effect calculations. Such also argues strongly against any threat of publication bias.

As for the primary studies included in this meta-analysis, the cohort designs that were typically used in the Title I assessments, even with external comparisons and mathematical adjustments, probably did not completely control for certain potential confound explanations. Regression toward the mean could be a particularly potent alternative explanation given the likely prevalence of extreme scores among bi-racial/ethnic samples (lower, on average, among African American children and higher among non-Hispanic white children). One can consider the probable direction of any such bias: Regression of extreme pre-program scores would likely cause the mean post-program scores among African American and non-Hispanic white children to be more similar than they had been (an apparent black-white achievement gap decrease) even without any programmatic effect. It seems that Title I interventions may be even less effective in narrowing the black-white test score gap than estimated by this meta-analysis. The quasi-experimental designs typical of the CSR assessments provided for fairly confident, internally valid inferences, though their

typical matching at the level of schools probably left some residual confounding on important individual and familial factors unaccounted for (e.g., parental socioeconomic status and household learning environment). However, this was a small group of studies, which in aggregate suffered in terms of external validity. Only one of this field's 12 widely implemented and well-researched CRS models was well represented—Success for All. Only two middle school and no senior high school outcomes were represented. And even though it was not a direct part of this meta-analytic review's search scheme, exhaustive searches did suggest that even fewer studies have examined the effects of Title I or CSR interventions on the underachievement of any other of America's potentially vulnerable racial/ethnic minority groups. This seems to be a particularly glaring knowledge gap given contemporary Title I-NCLB objectives.

Future Research Needs

Given the social significance of the black-white achievement gap as well as the great policy significance of educational programs that could effectively serve to ameliorate or eliminate it, investment in sound research toward these just ends would seem to be *the* moral imperative of our day. After synthesizing this field's limited knowledge base and reviewing the research methods that produced it, I can see no ethical rationale not to use experimentation in the future. What seems to be called for prior to investing further billions in educational innovations are more confident knowledge bases that could be produced by large, randomized controlled trials (RCTs).

This meta-analysis allows for the rather simplistic inference that “more is better,” but how much more funding is needed to transform an inequitable and ineffective traditional school into an equitable and effective one? The research community presently knows that such equitable and effective programs of instruction will be more “comprehensive” than is presently the case in traditional schools, but of the various theoretically important instructional elements of dozens of prevalent CSR models, which are most critical to success? A practical, inclusive, RCT-based, nation-wide research agenda across the elementary and secondary school continuums will undoubtedly be quite expensive, but such an investment holds the promise of huge educational and social dividends. It is time for the nation to make this commitment.

References

- * Studies included in the meta-analysis are preceded by an asterisk.
- * Berk, RJ. Doctoral dissertation. Vanderbilt University; 2003. Closing the black-white achievement gap: Instructional strategies in title I schools.
- Borman GD, D'Agostino JV. Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*. 1996; 18:309–326.
- Borman GD, Hewes GM, Overman LT, Brown S. Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*. 2003; 73:125–230.
- Braun HI, Wang A, Jenkins F, Weinbaum E. The black-white achievement gap: Do state policies matter? *Education Policy Analysis Archives*. 2006; 14(8) Retrieved March 4, 2008, from <http://epaa.asu.edu/epaa/v14n8/>.

- Brooks-Gunn J, Klebanov PK, Smith J, Duncan GJ, Lee K. The black-white test score gap in young children: Contributions of test and family characteristics. *Applied Developmental Science*. 2003; 7:239–252.
- Card, D., Rothstein, J. Racial segregation and the black-white test score gap. Princeton University, Industrial Relations Section; 2007.
- *. Centeno, BP. Doctoral dissertation. University of Southern California; 2005. Defeating the reading achievement gap at Fargo Elementary: To each according to his needs.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Erlbaum; 1988.
- Cooper, H. *Synthesizing research: A guide for literature reviews*. 3. Thousand Oaks, CA: Sage; 1998.
- *. Dawson, M. *Minority student performance: Is the Montessori magnet school effective?*. Washington: U.S. Department of Education, Office of Educational Research and Improvement; 1987.
- de Smidt GA, Gorey KM. Unpublished social work research: Systematic replication of a recent meta-analysis of published intervention effectiveness research. *Social Work Research*. 1997; 21:58–62.
- Downey DB, von Hippel PT, Broh BA. Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*. 2004; 69:613–635.
- Entwisle DR, Alexander KL. Winter setback: The racial composition of schools and learning to read. *American Sociological Review*. 1994; 59:446–460.
- Fleiss, JL., Levin, B., Paik, MC. *Statistical methods for rates and proportions*. 3. Hoboken, NJ: Wiley; 2003.
- Fryer RG Jr, Levitt SD. The black-white test score gap through third grade. *American Law and Economic Review*. 2006; 8:249–281.
- Fryer RG Jr, Levitt SD. Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*. 2004; 86:447–464.
- Glass GV, Smith ME. Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*. 1979; 1:2–16.
- Gorey KM. Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology Quarterly*. 2001; 16:9–30.
- Gorey KM, Cryns AG. Lack of racial differences in behavior: A quantitative replication of Rushton's (1988) review and an independent meta-analysis. *Personality and Individual Differences*. 1995; 19:345–353.
- Gorey KM, Luginaah IN, Schwartz KL, Fung KY, Balagurusamy M, Bartfay E, Wright FC, Anucha U, Parsons RR. Increased racial differences on breast cancer care and survival in America: Historical evidence consistent with a health insurance hypothesis, 1975–2001. *Breast Cancer Research and Treatment*. 2009; 113:595–600. [PubMed: 18330694]
- *. Gray, CS. Doctoral dissertation. Indiana State University; 2006. Effects of extended time programs on the academic achievement and social attendance of African American students and students in poverty in title I schools in the Evansville Vanderburgh School Corporation.
- Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews*. 1987; 9:1–30. [PubMed: 3678409]
- Greenwald R, Hedges LV, Laine RD. The effect of school resources on student achievement. *Review of Educational Research*. 1996; 66:361–396.
- Grenier AM, Gorey KM. The effectiveness of social work with older people and their families: A meta-analysis of conference proceedings. *Social Work Research*. 1998; 22:60–64.
- Grissmer, D., Flanagan, A., Williamson, S. Why did the black-white test score gap narrow in the 1970s and 1980s?. In: Jencks, C., Phillips, M., editors. *The black-white test score gap*. Washington, DC: The Brookings Institution; 1998. p. 182-226.
- Hedges LV, Laine RD, Greenwald R. Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*. 1994; 23:5–14.
- Hedges, LV., Nowell, A. Black-white test score convergence since 1965. In: Jencks, C., Phillips, M., editors. *The black-white test score gap*. Washington, DC: The Brookings Institution; 1998. p. 149-181.
- Hedges, LV., Olkin, I. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press; 1985.

- Herrnstein, RJ., Murray, C. *The bell curve: Intelligence and class structure in American life*. New York: Free Press; 1994.
- *. Hurley EA, Chamberlain A, Slavin RE, Madden NA. Effects of Success for All on TAAS reading: A Texas state-wide evaluation. *Phi Delta Kappan*. 2001; 82:750–756.
- *. Jones, BC. Doctoral dissertation. Tennessee State University; 2006. The effects of a basal reading program on reading achievement in selected Tennessee schools.
- Krueger, AB., Whitmore, DM. Would smaller classes help close the blackwhite achievement gap. Princeton University, Industrial Relations Section; Mar. 2001 (working paper #451)
- *. Ladewig, BG. Doctoral dissertation. University of Rochester; 2006. The minority achievement gap in New York State suburban schools since the implementation of NCLB.
- Lee, J. Tracking achievement gaps and assessing the impact of NCLB on the gaps: An indepth look into national and state reading and math outcome trends. Cambridge, MA: The Civil Rights Project at Harvard University; 2006.
- Levin HM, Belfield C, Muennig P, Rouse C. The public returns to public educational investments in African-American males. *Economics of Education Review*. 2007; 26:700–709.
- Magnuson KA, Duncan GJ. The role of family socioeconomic resources in the black-white test score gap among young children. *Developmental Review*. 2006; 26:365–399.
- *. Morgan, DH. Doctoral dissertation. Wilmington College; 1997. The relationship between a multi-age reading program in the primary grades and growth in reading achievement.
- *. Peasant, EJ, II. Doctoral dissertation. University of Southern Mississippi; 2006. Crossing the bridge from eight to tenth grade: Can ninth grade schools make it better?.
- Quint, J., Bloom, HS., Rebeck Black, A., Stephens, L., Akey, TM. The challenge of scaling up educational reform: Findings and lessons from First Things First. New York: Manpower Development Research Corporation; 2005.
- *. Ross SM, Smith LJ, Casey JP. “Bridging the gap:” The effects of the Success for All program on elementary school reading achievement as a function of student ethnicity and ability level. *School Effectiveness and School Improvement*. 1999; 10:129–150.
- Rothstein, R. *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. Washington, DC: Economic Policy Institute; 2004.
- Rushton JP, Jensen AR. Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*. 2005; 11:235–294.
- *. Shagle, S., Reynolds, M., Hoffer, TB. The effects of the Success for All program on student achievement in middle schools. National Opinion Research Center at the University of Chicago; 2008.
- *. Shepard, LK. Doctoral dissertation. Tennessee State University; 1996. A comparative study on academic achievement gains of Paideia students versus students attending a traditional school setting.
- Slavin, RE., Madden, NA. Reducing the gap: Success for All and the achievement of African-American and Latino students. Paper presented at the annual meeting of the American Educational Research Association; Seattle. 2001 Apr.
- Slavin RE, Madden NA. Reducing the gap: Success for All and the achievement of African-American students. 2006 Unpublished paper.
- Staines GL, Cleland CM. Bias in meta-analytic estimates of the absolute efficacy of psychotherapy. *Review of General Psychology*. 2007; 11:329–347.
- Torgerson CJ. Publication bias: The Achilles’ heel of systematic reviews? *British Journal of Educational Studies*. 2006; 54:89–102.

Biography

Kevin M. Gorey is the Assumption University-endowed Research Chair in Canadian and American Population Health. He is an epidemiologist interested in the affects of social policies on the health of diverse populations. This research review is related to the book he is

writing: “The skewed curve: Social solutions to America’s social problems.” It will be a “Bell curve” (Herrnstein & Murray, 1994) antithesis based on a series of planned systematic reviews of American education, social welfare and health care policies.

Table 1

Selected measures by program and school level

School level	Schools	Study outcomes	All participants	African Americans
<i>Title I program cohort observations</i>				
Elementary	91	8	5,842	2,685
Middle school	188	10	70,447	8,926
<i>Comprehensive School Reform quasi-experiments</i>				
Elementary	120	12	23,579	8,054
Middle school	16	4	3,596	2,336

Table 2

Meta-analysis results

<i>a. Title I program cohort observations</i>									
		Cohort initiation			Cohort follow-up				
	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃
Elementary schools	0.46	(0.41 0.52)	67.7	0.52	(0.46 0.58)	69.8			
Middle schools	0.32	(0.29 0.35)	62.5	0.27	(0.24 0.30)	60.6			
<i>b. Comprehensive School Reform quasi-experiments</i>									
		Traditional schools			CSR interventions				
	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃	<i>d</i>	95% <i>CI</i>	<i>U</i> ₃
Elementary schools	0.64*	(0.61 0.67)	73.9	0.15	(0.12 0.18)	55.9			
Middle schools	0.21	(0.14 0.28)	58.3	0.06	(-0.01 0.13)	52.4			

* Distribution significantly heterogeneous, *p* < .001.