

**National Efforts to Bring Reform to Scale in America's High-Poverty
Elementary and Secondary Schools: Outcomes and Implications**

Geoffrey D. Borman

University of Wisconsin-Madison

Paper commissioned by the Center on Education Policy, Washington, D.C.
For its project on Rethinking the Federal Role in Education

March 2009

The views expressed in this paper are those of the author.

Abstract

Since the 1960s, there have been continuing federal efforts to bring reform to scale in high-poverty elementary and secondary schools across the U.S. This paper traces the evolution of these efforts and discusses their impacts on achievement outcomes. Drawing on evidence from meta-analyses of the Title I evaluation literature and the Comprehensive School Reform research base, four general themes emerge.

First, there has been a clear developmental trajectory of these efforts from 1965 to the present that has resulted in historical improvements in disadvantaged students' outcomes. Second, although the achievement effects have been somewhat modest, the evidence suggests that these national efforts are capable of contributing to large-scale improvements in high-poverty schools. Third, there is great variability across schools and time in the outcomes of these reform efforts that can be explained by both the level of implementation of reforms and by differences in the methods researchers have used to estimate their effects. Fourth, high schools have historically been underserved by federal policies to reform high-poverty schools, but growing interest among policymakers and accumulating evidence related to "relevance," "rigor," and "relationships" might help direct future investments toward improving America's high schools. A number of promising models for reforming the nation's high-poverty schools exist, but higher quality studies and better standards of evidence are needed to help advance the scale-up of scientifically based interventions.

National Efforts to Reform Elementary and Secondary Schools

Education in the United States is a decentralized system comprised of highly variable practices, programs, and school contexts. The primary technology of education—teaching—is highly complex and is typically designed and implemented by teachers who have traditionally enjoyed a great deal of autonomy and independence from regular inspection. The principal goals and products of education—and whether they should be ones centered around creativity, knowledge of basic facts, sound moral judgment, or something else—are constantly open to differing opinions and debate. Can such a diffuse and decentralized system with uncertain technology and goals be served by centralized efforts to implement educational reform at scale? Further, how can educational research support the scale-up of promising programs and practices?

This article discusses some of the ways in which recent national efforts to reform the country's elementary and secondary schools inform these questions. I begin by tracing the recent history of these reform efforts. In addition to considering the evolution of the interventions themselves, I describe how the national research and evaluation agenda has evolved, and can continue to evolve, to help advance the development of replicable programs and evidence-based educational policy. Finally, applying the lessons learned from two syntheses of the federal Title I and comprehensive school reform research literature and from more recent research on replicable programs and practices for reforming American elementary and high schools, I offer four conclusions for methodologists, policy makers, and reform developers to consider when conducting

research, crafting policy, and refining educational programs to support the scale-up of educational innovation.

The Recent History of National Reform Efforts

Since the advent of a national effort to improve the United States' most challenged high-poverty elementary and secondary schools, the capacity, technology, and policy to support the scale-up of school reform has expanded dramatically. The roots of this movement can be traced back to 1965 when Title I of the Elementary and Secondary Education Act (ESEA) was implemented as a centerpiece of Lyndon B. Johnson's War on Poverty, "to provide financial assistance to...local educational agencies serving areas with concentrations of children from low-income families to expand and improve their educational programs by various means...which contribute particularly to meeting the special educational needs of educationally deprived children" (ESEA of 1965, 79 Stat. 27, 27). Along with the emerging system of social programs of the 1960s, Title I was the major educational initiative designed to close the achievement gap between poor children and their more advantaged peers, and, ultimately, to break the vicious cycle of poverty.

Problems of Implementation

In the early years, the nation's efforts to bring reform to scale in high-poverty schools were largely sabotaged by ineffectual policy and a nearly non-existent knowledge base of how to improve schools for the disadvantaged. McLaughlin (1976) noted that the original program mandates were ambiguous concerning the proper and improper uses of Title I funds, and the guidelines and intent of the law were open to interpretation.

Varying local interpretations of the law, rather than clear and uniform federal mandates, guided the use of the federal funds.

Also, in 1965, the research base and practitioner knowledge base for developing effective compensatory education programs were extremely limited. The majority of local administrators and teachers lacked the experience and understanding for developing, implementing, and teaching compensatory programs. Though research provided some basic descriptions of exemplary practices in select sites (Hawkrige, Campeau, DeWitt, & Trickett, 1969; Hawkrige, Chalupsky, & Roberts, 1968; Wargo, Campeau, & Tallmadge, 1971), there were no clear replicable programs that could be scaled up to serve large numbers of schools.

Although the federal dollars provided localities an incentive to improve education for the disadvantaged, a viable intergovernmental compliance system was not in place. Without effective regulation, the receipt of funds did not depend on meeting the letter or the spirit of the law. Responding to local self-interests and utilizing Title I dollars for established general aid policies was an easier option than the new and more complicated task of implementing effective programs for poor, low-achieving students.

Despite early resistance by most federal policymakers to restrict local control, the continued misuse of Title I funds by various states, districts, and schools along with growing pressures exerted by local poverty and community action groups prompted the U.S. Office of Education to reconsider the legislative and administrative structure of Title I (Jeffrey, 1978; Kirst & Jung, 1982). During the 1970s, the Congress and U.S. Office of Education established more prescriptive regulations related to school and student selection for services, the specific content of programs, and program evaluation, among

other things (Herrington & Orland, 1992). These additional responsibilities placed greater administrative demands on local school systems. Funded in part by federal dollars, larger and more specialized state and district bureaucracies grew to monitor local compliance. State and local compliance was confirmed through periodic site visits and program audits by the U.S. Office of Education and by the Department of Health, Education, and Welfare. As Cohen (1982) and Meyer, Scott, and Strang (1986) noted, the Title I legislation of the 1970s, along with the proliferation of other state and federal educational mandates, promoted the expansion and increased bureaucratization of local educational agencies.

Reform through Bureaucratization

As the 1970s progressed, the bureaucratic organization of Title I became institutionalized across the country and services were delivered to the children targeted by the law (Peterson, Rabe, & Wong, 1986). Rather than a heavy federal presence with intergovernmental conflict, the implementation of Title I became a cooperative concern and professional responsibility of local, state, and federal administrators. In addition, Peterson et al. noted that Title I had inspired greater local concern for, and attention to, the educational needs of the children of poverty. Therefore, in marked contrast to the first decade of the program, during the latter half of the 1970s and throughout the 1980s the specific legislative intents, and the desired hortatory effects, were achieved on a far more consistent basis.

Though the program was reaching the students it had targeted during this era, the actual practices were driven more by bureaucratic regulations than by any research-based

or practitioner-developed model of what constituted effective education services for disadvantaged children. One of the most important regulations affecting program delivery had been the provision that the compensatory services provided through Title I must supplement, not supplant, the regular educational programs provided to eligible students. In case of program audits, and to clearly account for the federal money, educators and administrators needed to show that the targeted Title I programs actually provided something “extra,” and that they were not merely replacing services that the students would have received through the regular school program.

This regulation led to widespread use of the “pullout model” as a means for delivering supplemental compensatory services to eligible Title I students. Most often, the students who qualified for services were taken, or “pulled out,” of their regular classrooms for 30 to 40 minutes of remedial instruction in reading and math. This arrangement had the advantage of making it clear that the funds were providing something separate from the regular school program, as special teachers, books, and other materials were clearly allocated only to the pulled-out Title I students and not their regular classroom peers. Despite some research suggesting that pullout programs stigmatized children and provided few, if any, academic benefits (Glass & Smith, 1977), through the 1970s, 1980s, and much of the 1990s about three of four Title I schools used the pullout model to deliver supplemental services.

Combining Flexibility with Accountability for Improvement

Instead of the seemingly piecemeal and uncoordinated categorical targeted assistance programs that had served Title I schools since the mid-1960s, a growing belief

developed that at-risk students and high-poverty schools could be better served by schoolwide reforms. This belief was encouraged by informed opinion (e.g., Rotberg, Harvey, & Warner, 1993), by general findings from the effective schools research tradition (Edmonds, 1979; Teddlie & Reynolds, 2000), and by the concept of systemic reform (e.g., Smith & O'Day, 1991), more than by specific groundbreaking empirical studies. Inspired by the emerging vision of standards-based reform, the 1994 reauthorization of Title I called on states to raise academic standards, to build the capacity of teachers and schools, to develop challenging new assessments, to ensure school and district accountability, to ensure the inclusion of all children, and to develop coordinated systemic reforms. The new legislation encouraged schoolwide initiatives rather than targeted programs for all schools where at least 50% of the students were poor. The new Title I legislation encouraged schools to use the funds with greater flexibility to support ongoing school-based reform efforts or initiate new ones to help address the educational needs of all children from high-poverty schools. These sweeping changes began the transformation of Title I from a supplemental remedial program to the key driver of the standards-based, schoolwide reform movement (Borman, 2000).

During the 1990s, Title I schoolwide projects proliferated across the country. In 1991, only 10% of the eligible Title I schools operated schoolwide programs, but by 1996, approximately 50% of the eligible Title I schools had implemented them (Wong & Meyer, 1998). A number of studies from the 1990s showed that, in the short-term, these schoolwide efforts did not produce compelling evidence of positive achievement effects and, for the most part, did not result in the desired reforms (Wong & Meyer, 1998, 2001). Also during the 1990s, a more general review indicated that site-based management reforms failed to

affect student outcomes positively in large part because the schools failed to develop coherent statements of beliefs or models for guiding the work and decision-making of the school (Murphy & Beck, 1995). These outcomes, combined with new evidence from the Congressionally mandated Prospects study of the modest overall impacts of Title I services (Borman, D'Agostino, Wong, & Hedges, 1998; Puma et al., 1997), suggested that federal policies for improving education for at-risk students from high-poverty schools were in need of further retooling. Despite the new flexibility afforded by the law, the largely locally inspired, schoolwide reforms did not yield the desired effects on educational practices and outcomes.

At the same time, the growing research base on several externally developed school restructuring efforts, such as the Comer School Development Program (Comer, 1988; Haynes, Emmons, & Woodruff, 1998) and Success for All (Slavin & Madden, 2001), seemed to indicate hope for a high-quality education for at-risk students. In addition, the companion study to the national Prospects evaluation of Title I, the Special Strategies Study (Stringfield et al., 1997), indicated that whole-school, externally developed programs funded by Title I appeared more likely to have positive impacts on academic achievement than either traditional Title I pullout programs or locally developed school reform efforts.

Scaling up Reform with Evidence-Based Replicable Programs

Along with growing policy and research support, in 1991 then-President George Bush announced the creation of a private-sector organization called the New American Schools Development Corporation (NAS), which was intended to support the creation of “break the mold” whole-school restructuring models for the next century (Kearns &

Anderson, 1996). Using a business model, NAS turned to the marketplace for proposals for new models of American schools that would enable all students to achieve world-class standards in core academic subjects, operate at costs comparable to current schools after start-up funding, and address all aspects of a school's operation. After receiving nearly 700 proposals in February 1992, NAS chose 11, and provided funds for a three-year program of development and testing. From 1995 through 2004, NAS continued to focus on scaling up seven of the models to thousands of schools nationwide. Providing more than \$150 million in financial and technical assistance to the reform developers, NAS helped create a market for comprehensive school reform (CSR) and helped scale up the CSR movement.

In response to the promise of the externally developed programs disseminated by NAS and by other independent model developers, the U.S. Congress also has encouraged individual schools to implement "scientifically based" whole-school reforms and to seek the assistance of external groups in developing their school reform plans. In 1998, Congress initiated the Comprehensive School Reform Program (CSRP), which encouraged schools to develop comprehensive plans for implementing "scientifically based" strategies for school reform. Through a competitive process, CSRP awarded a minimum of \$50,000 per year for three years to qualifying schools. Since first authorizing CSRP in fiscal year 1998 and allocating a total of \$145 million, Congress steadily increased its support. In fiscal year 2002, allocations for the CSRP equaled \$310 million. This figure included \$235 million set aside specifically for Title I schools and \$75 million available to schools wishing to apply through the Fund for the Improvement of Education.

The other significant funding source for CSR programs has been Title I. In January 2002, with the reauthorization of Title I as the No Child Left Behind Act (NCLB), the CSRP and Title I came together under the same legislation. As Title I, Part F, CSRP has become a significant component of the growing federal movement to support scientifically based efforts to reform low-performing high-poverty schools across the nation. This federal support, combined with the efforts of NAS and other independent developers, led to an expansion of externally developed, evidence-based school reform models.

Though federal funding for CSR has dwindled in recent years, from the early 1990s through the early 2000s, the scale-up of evidence-based CSR designs happened at an unprecedented rate, as evidenced by the growing number of externally developed school reform designs (e.g., Accelerated Schools, Core Knowledge, High Schools That Work, Success for All) that were implemented in thousands of schools, serving millions of students throughout the United States. CSR focuses on reorganizing and revitalizing entire schools, rather than on implementing a number of specialized, and potentially uncoordinated, school improvement initiatives. In general, the funding sources supporting the implementation of CSR have been targeted toward the schools most in need of reform and improvement: high-poverty schools with low student test scores. According to 1998-2006 data from the Southwest Educational Development Laboratory, schools receiving money to implement CSR models through the CSRP have an average poverty rate of 70%. Further, nearly 40% of schools receiving CSRP funds were identified for school

improvement under Title I regulations and over 25% were identified as low-performing schools by state or local policies.¹

Some schools develop their own “home-grown” reform models, but many educators are turning to groups external to the schools, such as universities and educational centers and labs, for assistance in designing whole-school reform models. Externally developed reform designs are consistent in that they provide a coherent schoolwide model for instructional and organizational change. At the same time, though, the externally developed designs are remarkably diverse in their analyses of the specific problems in U.S. education, the solutions that they propose, and the processes through which they propose that schools may achieve those solutions.

The Comer School Development Program, for example, builds largely around Dr. James Comer’s work in community psychiatry and focuses its energy on creating schools that address a wide range of students’ health, social, emotional, and academic challenges. By contrast, the Success for All program (Slavin & Madden, 2001) offers a well-specified school reform model that focuses primarily on prevention of reading difficulties during the early elementary school years. The Coalition of Essential Schools model attempts to create more educationally rich and supportive learning environments through a common adherence to nine broadly philosophical, common principles (Sizer, 1992), whereas the Talent Development High School (Letgers, Balfanz, Jordon, & McPartland, 2002) features highly specified components including a self-contained 9th grade academy,

¹ This information was obtained from the Southwest Educational Development Laboratory’s CSRD database, which is available online at <http://www.sedl.org/expertise/historical/csr-awards-database.html>. The data reported here include all schools receiving CSRP awards that began in 1998, 1999, 2000, and 2001. Not all schools reported whether they had been identified for improvement under Title I, state, or local regulations. Therefore, the percentages reported are, most likely, underestimates.

career academies for grades 10-12, and extra help for students delivered through the Twilight School.

CSR expanded rapidly because many models established development and dissemination infrastructures for replicating and supporting implementations across numerous schools. In other words, the developers can transport their CSR models to schools across the U.S., help local educators understand the tenets of the reform, and teach them how to implement the school organization and classroom instruction that the model suggests. In every case, the developers provide some type of initial training or orientation to help educators to at least understand the underlying philosophy of the model. In many circumstances, replication also involves a more specific “blueprint” for implementing and sustaining the model. Highly specified models, for instance, often prescribe new curricular materials, new methods of instruction, alternative staffing configurations, and a series of ongoing professional development activities.

Along with CSR, the federal government further emphasized the scale-up of replicable research-proven programs through other important initiatives. For instance, the federal Reading First program has offered states and districts support to apply scientifically based reading research—and the proven instructional and assessment tools consistent with this research—to ensure that all children learn to read well by the end of third grade. Like CSR, states provide subgrants to eligible districts on a competitive basis. The program asks state education agencies to fund those proposals that show the most promise for raising student achievement and for successful implementation of reading instruction, particularly at the classroom level. Only programs that are founded on scientifically based reading research are eligible for funding through Reading First.

The education research and development infrastructure supporting the scale-up of research-proven programs also has received significant upgrades in recent years.

Established by the Education Sciences Reform Act of 2002 and under the leadership of Grover (Russ) Whitehurst, the Institute of Education Sciences (the research arm of the U.S. Department of Education) led a prominent nationwide push to promote the use of randomized experiments for evidence-based decision making (Whitehurst, 2002).

Since 2002, new grant competitions designed by IES have focused on the development of practical solutions to improve public schools in the U.S. and have emphasized application of high-quality methods of causal inference including, when possible, randomized designs. Also, in 2006, IES continued the federal commitment to educational research and development by funding 10 regional educational laboratories committed to providing policymakers and practitioners with expert advice, training, and technical assistance on how to interpret the latest findings from scientifically valid research pertaining to the requirements of No Child Left Behind (Bowler & Thomas, 2006) in instances where scientific evidence is not readily available and schools need appraisals of alternative strategies to improve learning, IES charged the laboratories to devote approximately one-third of their operating budgets to carrying out rigorous randomized trials to evaluate potentially promising practices and programs.

Finally, the reauthorization of the ESEA of 1965 as the No Child Left Behind Act of 2001 required practices based on high-quality research for everything from the technical assistance provided to schools to the choice of anti-drug-abuse programs. Within the No Child Left Behind Act, phrases like “scientifically based research” appear more than 100 times (Olson & Viadero, 2002). Like CSR, Reading First, and other recent

federal education programs, this legislation also places a premium on randomized experiments for developing and assessing new and innovative practices, as the following excerpt suggests: “The Secretary shall evaluate the demonstration projects supported under this title, using rigorous methodological designs and techniques, including control groups and random assignment, to the extent feasible, to produce reliable evidence of effectiveness” (No Child Left Behind Act of 2001, 115 Stat 1425, 1597). This legislation, urging the use of scientifically based educational methods and procedures, is meant to revolutionize not only the cornerstone of the ESEA, Title I schoolwide and targeted assistance programs for the disadvantaged, but also the Reading First and Early Reading First programs, the Even Start family literacy programs, services for limited English proficient students, and other federal initiatives.

Also of significance, the NCLB legislation further developed and expanded the previous accountability mandates of Title I and asked all states to develop achievement tests to hold schools accountable across grades 3 through 8 and one high-school grade level. The new law substantially increased the Title I testing requirements and set very demanding accountability standards for schools, districts, and states, including setting measurable adequate yearly progress objectives for all students, as well as for subgroups of students defined by socioeconomic background, race/ethnicity, and English language proficiency. Schools are required to demonstrate that every subgroup of students meets adequate yearly progress (AYP) targets for both participation and proficiency in mathematics and literacy. NCLB also bolsters the consequences associated with consecutive years of AYP failure. Schools that miss AYP targets for two consecutive years are identified for improvement and must offer Title I choice. Those that fail three

consecutive years must offer supplemental educational services. Failure to meet AYP targets for four or more consecutive years results in designations of corrective action and restructuring, for which the sanctions stiffen each subsequent year. Though test-based accountability has been an enduring feature of Title I since the advent of the Title I Evaluation and Reporting System (TIERS) during the 1970s, these new efforts placed even stronger mechanisms in place to focus the attention of educators and policymakers on specified AYP targets and to provide stronger accountability in the form of rewards and sanctions related to schools' progress toward AYP.

Four Stages of Development in the National Reform Movement

This series of initiatives in the national movement to bring reform to scale in high-poverty elementary and secondary schools has a clear developmental trajectory that can be summarized by four distinct stages. First, the early implementation of Title I was characterized by intergovernmental conflict, poor implementation, and a lack of research-based and practitioner-based knowledge of how to develop effective educational interventions for disadvantaged students. A second stage, during the 1970s and 1980s, was marked by the development of increasingly specific policies to guide the Title I program's implementation and evaluation, growing bureaucratic cooperation between federal and local authorities in implementing the policies, and improved access for disadvantaged students to the supplemental resources and instruction offered by the program.

Rather than simple access to supplemental services, during the late 1980s and 1990s new Title I legislation stressed reform and improvement of the program. The

emphasis on emerging national education standards and systemic reform supplanted many of the earlier concerns about fiscal and procedural accountability, as this latter type of accountability was all but taken for granted. In keeping with the national trends toward site-based decision making and decentralization, Title I afforded schools greater flexibility to serve disadvantaged students, so long as their test scores improved. For the most part, though, this flexibility did not prompt schools to develop new visions for reform. Aside from some tinkering around the edges, the administration and operation of Title I remained fairly stable.

Beginning in the 1990s, the current stage emerged in which the scale-up of research-proven programs and practices has been increasingly regarded as the key to improving the effectiveness of high-poverty elementary and secondary schools. Like the 1980s and 1990s, the general spirit of today's reform efforts continues to articulate top-down education standards and even stronger test-based accountability mechanisms, which dictate many of the changes in the content of schooling. However, the process of reform and the mechanisms to improve instruction and build school capacity are in marked contrast to the earlier stages of Title I. Rather than policy mandates or flexibility alone, a growing constellation of replicable programs has become a key lever through which educational practices and the processes of school change may be shaped.

In many ways, this recent focus on replicable programs helps reconcile the two most important recent educational reform movements in the United States. Since the 1980s, competing and often contradictory reforms have combined top-down, centralized efforts to improve schools and teaching with efforts at decentralization and school-based management (Rowan, 1990). The problem is that the complex educational changes

demanded by current standards-based reform initiatives, combined with an increasingly heterogeneous student population largely composed of students whom schools have traditionally failed, have pushed the technology of schooling toward unprecedented levels of complexity. In many ways, expecting local educators to reinvent the process of educational reform school by school is both unrealistic and unfair. Externally developed educational programs provide a type of top-down direction for designing and supporting the process of school reform. In this case, however, the top-down direction is not in the form of distant legislative mandates, but is, in theory, tangible and accessible support for school improvement rooted in research and literally packaged and delivered to each school's door.

Evidence of Effects on Achievement Outcomes

Given the apparent progress made in scaling up reform in high-poverty elementary and secondary schools, it should come as little surprise that recent evidence suggests that these efforts to meet the needs of disadvantaged children have helped the United States make strides toward greater educational equality. The long-term trend data from the National Assessment of Educational Progress (NAEP) indicated tremendous progress beginning in the 1970s and 1980s in closing the persistent achievement gaps separating poor and more advantaged children and African American and white students (Grissmer, Flanagan, & Williamson, 1998; Smith & O'Day, 1991). For instance, during this period the gaps between African American and white children shrank by about two grade levels. The reasons for this unprecedented trend are open to some debate, but Grissmer and his colleagues asserted that Title I and the other social and educational

programs that were first introduced during the War on Poverty of the mid-1960s surely had something to do with it.

A Meta-Analysis of Title I Effects

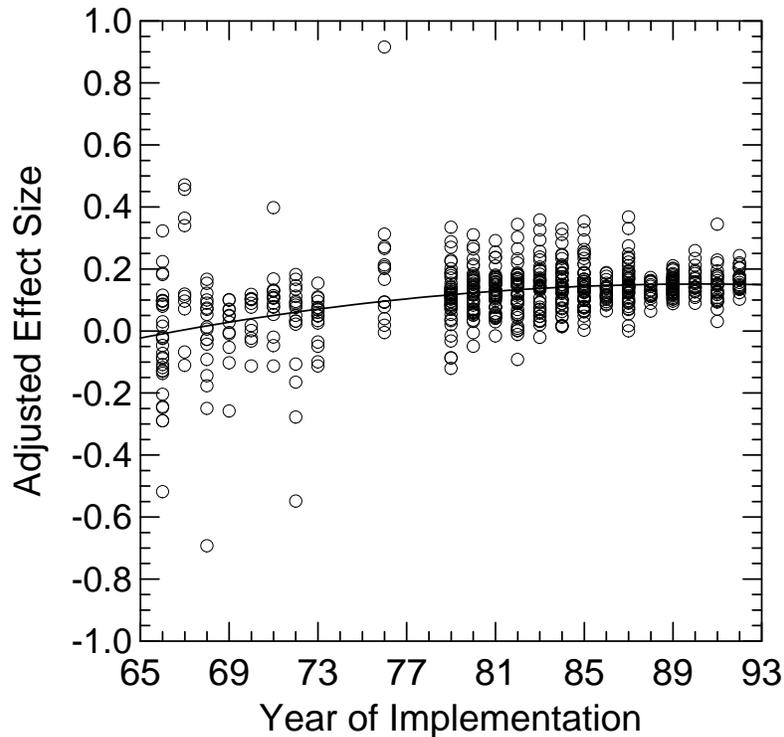
Supporting this assertion, a comprehensive meta-analysis, or quantitative review, of the results from 17 federal evaluations from 1966 through 1993 indicated that the 1970s and early 1980s were also the periods of the greatest improvements in Title I students' math and reading achievement outcomes (Borman & D'Agostino, 1996; 2001). During the early years of Title I, in the late 1960s, the program was not effective in closing the gap because it simply was not implemented as intended by Congress. As the regulations and knowledge base for implementing Title I programs came into clearer focus during the 1970s and 1980s, the intended recipients of the program's services, largely poor and African American children, began to show clear benefits from Title I, and the nation's achievement gaps began to close.

Although it is not possible to establish a true cause-effect relationship between the closing gaps and the improvements in Title I students' outcomes, two points are clear. First, Borman and D'Agostino's meta-analysis suggests that the children served by Title I would have been worse off academically without the program. Second, the fact that important national progress was made in closing the achievement gaps demonstrates that educational inequality can be overcome and potentially eliminated in a relatively short period of time. Third, the fact that significant new policies and funding sources—the War on Poverty programs, and ESEA most notably—were specifically targeted toward improving education and other services for disadvantaged children and their families

during this time suggests that they are likely to have played a role in this improvement. Indeed, these outcomes suggest that the scale-up of programs for high-poverty schools can help contribute to widespread effects on student outcomes.

Beginning in the late 1980s, however, the important gains made by African American and poor children began to slow and even erode somewhat (Grissmer et al., 1998). Once Title I was effectively implemented as intended by Congress during the late 1970s and early 1980s, the promising gains made by participating children also plateaued (Borman & D'Agostino, 2001). After statistically taking into account a variety of programmatic and methodological moderators that have influenced the estimates generated by national evaluations of the Title I effect size over the years 1965 through 1994, Borman and D'Agostino (2001) obtained the residuals from the regression. By fitting the average Title I effect size of $d = .11$ to each residual, the resulting scatterplot of adjusted effect size by year of implementation displayed in **Figure 1** provides a visual representation of how Title I effects have changed over the years, after taking into account the differences across the evaluations of the program.

Figure 1. Adjusted Effect Size by Year of Title I Implementation



The figure plots the adjusted Title I effect sizes by the year of implementation, 1965 through 1993. The line of best fit through the plotted effect sizes shows a trend of increasing Title I effects from 1966 through the early 1980s and a plateau in the effects reached during the 1980s and 1990s.

Source: Borman, G.D., & D'Agostino, J.V. (2001). Title I and student achievement: A quantitative synthesis. In G.D. Borman, S.C. Stringfield, & R.E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 25-58). Mahwah, NJ: Lawrence Erlbaum Associates.

The figure contains 657 data points, each representing an independent estimate of the Title I effect derived from 17 national studies and including the test scores of over 41

million Title I students from grades 1 through 12. The line of best fit through the data points indicates a somewhat nonlinear relationship between adjusted effect size and year of implementation. Specifically, Figure 1 shows a linear improvement in program effects from 1966 to the early 1980s, increasing from an effect size of about 0 in 1966 to an effect of nearly .15 in the early 1980s. This suggests that when localities implemented programs of variable, but generally poor, quality during the 1960s, the effects were, on average, essentially zero. Improved implementation led to improvements in the effectiveness of the program during the 1970s. However, beginning in the 1980s, the effects plateaued, remaining at around .15 throughout most of the 1980s and the early 1990s.

This pattern of improvement in Title I effects suggests that once the program was implemented as intended by Congress during the late 1970s and early 1980s, the effects reached a peak that has not changed substantially. The pattern of variability in program effects also supports this conclusion. The wide variation in program effects during the 1960s and early 1970s appears to reflect the variability of local program implementation and evaluation. However, once implementation and accountability requirements became more uniform and established throughout the late 1970s and 1980s, this not only led to increased effectiveness, but to more consistent effectiveness. One might conclude that this result suggests that an effect of 0.15 is the best we can do given the current federal funding commitment and structure of the program. Alternatively, it could be taken as a sign that the standardized, and modestly effective, procedures of Title I's more recent history require substantial reform in order to promote continued improvement.

A Meta-Analysis of CSR Effects

With the No Child Left Behind Act and CSR, this reform movement gained significant momentum during the late 1990s and early 2000s. Though this movement has slowed, the more general idea of research evidence driving the development and dissemination of research-proven educational programs has continued to thrive. The meta-analysis by Borman, Hewes, Overman, & Brown (2003), synthesized evidence regarding the achievement effects of 29 widely replicated CSR models. The 29 models selected for the research synthesis were implemented in 55.6% of the schools that received CSRP funds for externally developed models, as reported in the Southwest Educational Development Laboratory (SEDL) database. Therefore, the results of the review generalize reasonably well to the population of U.S. elementary and high schools implementing CSR models using CSRP and Title I program funds.

So how do CSR effects compare to the previous national efforts to help close the achievement gap and improve the outcomes of large numbers of high-poverty and low-achieving students and schools? The most obvious comparison to the effect of CSR programs is the effect of the traditional Title I programs that preceded them, which were the subject of Borman and D'Agostino's (1996) earlier meta-analysis. The overall mean weighted effect size of CSR of $d = .15$ compares favorably to the overall average weighted Title I effect of $d = .11$, but because the primary studies and the two meta-analyses used somewhat different methodologies, the comparison is imperfect.

A better comparison between CSR and conventional Title I programs may be drawn directly from the Borman et al. (2003) meta-analysis by examining the CSR effect sizes estimated from the comparison-group studies in schools with 50% poverty or more.

In most of these cases, the comparison schools had such high poverty rates that it was highly likely that they received federal Title I funds. In most cases, these schools implemented Title I targeted or schoolwide programs and were not implementing other CSR models. These studies, therefore, provided a relatively good indication of the value-added effects of CSR, above and beyond the effect of traditional Title I programs. Across 346 such comparisons, the effect size, statistically adjusted for methodological characteristics, was $d = .12$. In other words, despite the fact that the vast majority of these control schools provided their students with extra resources and programs provided through Title I, the average CSR school still outperformed 55% of the Title I schools.

Drawing on national evidence from NAEP and from meta-analytic estimates of the effects of Title I and CSR, at least two points come into clearer focus. First, there appears to be national progress in scaling up improved educational outcomes for students and schools from disadvantaged circumstances. This is marked by progress in closing the achievement gaps separating African Americans and whites and poor and non-poor students. It is also distinguished by the trend of growing achievement effects associated with national efforts to reform high-poverty schools through Title I, CSRP, and other evidence-based programs and practices. These outcomes suggest that national efforts to scale up reform in high-poverty schools are capable of producing widespread improvements in educational outcomes. In the aggregate, though, these national effects are somewhat modest. They amount to no more than effect sizes ranging from $d = .11$ to $d = .15$.² However, as suggested by the great variability in schooling across the diverse contexts in which it is carried out, the variation in the effects of scaling up reform are

² These achievement effects are also fairly consistent with experimental estimates from the recent Tennessee STAR study of the educational outcomes of the statewide scale-up of reform through reductions in class size (Finn & Achilles, 1999).

often a more significant part of the story than the aggregate effects.

Explaining the Variability of Effects

Perhaps the most salient theme of the meta-analyses of Title I and CSR research is that the overall effects of these national efforts to bring reform to the nation's high-poverty elementary and secondary schools are marked by considerable heterogeneity. Rather than a distinct and replicable model for reform, Title I is better understood as a funding mechanism that allows for extensive variation, both across and within schools, in design and implementation. Some schools operate Title I programs that serve all students schoolwide, whereas others operate programs that target only the lowest-achieving students within the school. Some schools may also, for example, spend all of their Title I funds on helping 9th graders learn basic math skills, but other schools may channel their resources toward helping students across the grades master literacy skills. As a consequence, and as the results from Borman and D'Agostino's (1996, 2001) meta-analytic work suggest, any overall "treatment effect" is best viewed as random rather than fixed, in that a single estimate of the population effect for Title I is not likely to generalize across schools and programs.

Across the 29 CSR models, as one might expect, there is also a considerable amount of variability in their effects. As one might also expect, there is of course less variability across schools implementing any one of the 29 models because, in contrast to Title I, each of the 29 models offers a relatively distinctive and replicable model for school reform. There are also a number of discrete features of CSR programs, either

called for by the U.S. Department of Education or those that have been the topic of prior research that one may identify as key ingredients of reform across the 29 models.

In attempting to evaluate empirically how various reform model components helped us statistically account for differences among schools in their achievement outcomes, however, we found that they told us very little. In fact, whether or not the various reform models called for ongoing staff professional development, measurable goals and benchmarks for student learning, a faculty vote to increase the likelihood of model acceptance and buy-in; and the use of specific and innovative curricular materials and instructional practices designed to improve teaching and student learning had little bearing on the achievement outcomes the schools produced.. Similarly, the frequency with which the CSR models have successfully replicated their approaches in schools with diverse characteristics, the overall level of external technical support and assistance from the developer, and the general cost of the model do not help to explain a substantial amount of the variability in the CSR effect across schools.

The fact that the school reform components provided so little insight into school-to-school differences in their achievement outcomes suggests at least two possible interpretations. The first is that these components are not important for promoting student achievement in CSR schools and, therefore, there is no relationship. The second interpretation is that knowing whether or not a CSR model generally required schools to implement a given component tells us little about whether or not the component actually was implemented. This latter interpretation suggests that some or all of these components may make a difference in terms of student achievement, but school-specific and model-specific differences in the ways that the components are actually implemented explain

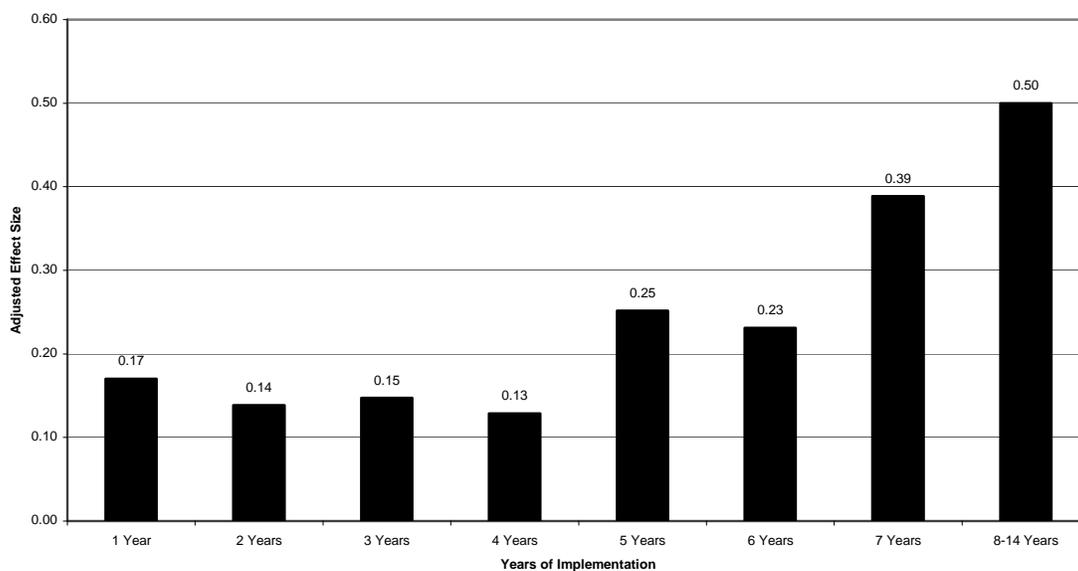
considerably more than simply knowing whether or not the CSR developer requires them. Prior research has linked the success of school reform to the level and quality of implementation (Berman & McLaughlin, 1978; Crandall et al., 1982; Datnow, Borman, & Stringfield, 2000; Stringfield et al., 1997), the coordination and fit of the model to local circumstances, and the relationship between the CSR developer and the local school and school district (Datnow & Stringfield, 2000). Knowledge of these factors, which have been largely unmeasured and unreported in evaluations of the achievement effects of CSR programs, would enrich our understanding of the variability in the CSR effects.

Indeed, with respect to the variability of outcomes found for both Title I and CSR, one of the most convincing findings from both meta-analyses is simply that implementation matters. The history of Title I has shown a strong relationship between implementation and program effects, as measured by students' achievement outcomes. Similarly, the best available measure of level of implementation from the meta-analysis of CSR research, the number of years that a CSR model was implemented at a school, shows a similar outcome. **Figure 2**, which combines evidence from across the 29 CSR models, displays effect sizes by the number of years of CSR program implementation. The finding across the 29 models is consistent in showing an increasing effect on achievement outcomes associated with a greater number of years of implementation.

The figure shows that the CSR effect size, .17, was relatively strong during the first year of implementation. Then, perhaps reflecting the “implementation dip” that Fullan (1991) has noted from his conversations with principals and teachers, there appears to be a tendency for new CSR initiatives to get somewhat worse before they get better. This is reflected by the slight decline in effect sizes during the second, third, and

fourth years of implementation. After the fifth year of implementation, however, the CSR effects began to increase substantially. Schools that had implemented CSR models for five years showed achievement advantages that were nearly twice those found for CSR schools in general, and after seven years of implementation, the effects were more than two and half times the magnitude of the overall CSR impact of $d = .15$. The small number of schools that had outcome data after 8 to 14 years of CSR model implementation achieved effects that were three and a third times larger than the overall CSR effect.

Figure 2. Adjusted Effect Size by the Number of Years of CSR Model Implementation



The figure plots the adjusted CSR effect size by the number of years that the model was implemented at the school. Models that had been implemented for 5 years or more showed the most substantial impacts on achievement.

Source: Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73, 125–230.

These strong effects of CSR that begin after the fifth year of implementation may be explained in two ways: a potential cumulative impact of CSR or a self-selection artifact. Specifically, schools may be experiencing stronger effects as they continue implementing the models, or it could be that the schools experiencing particular success continue implementing the reforms while the schools not experiencing as much success drop them after the first few years. Both explanations are plausible. Nonetheless, it is of considerable significance that the average school across all studies reviewed in the meta-analysis had implemented its CSR model for approximately three years. The typical study in the meta-analysis, therefore, may have underestimated the true potential of CSR for affecting change in schools and for improving student achievement.

The Special Case of High Schools

Although education reform has been a prominent national issue since the release of *A Nation at Risk* in 1983, most of the attention has been focused on elementary grades and improving basic reading and math skills for younger students. Recent federal initiatives, including CSRP, The No Child Left Behind Act, and Reading First target the vast majority of attention and funding to reforms in the early grades. Only the Carl D. Perkins Vocational and Technology Education Act has played any significant role in providing resources to high schools.

Though No Child Left Behind does hold high schools and school districts accountable for high school graduation rates and student performance on one high school assessment, the clear strength of the legislation is targeted toward accountability for grades 3 through 8. Indeed, the widely held belief that Title I has a more profound impact

in the elementary rather than the high school grades is supported by the distribution of effect sizes across grades from Borman and D'Agostino's (1996) meta-analysis of Title I effects. Further, though the meta-analysis of school reform models by Borman et al. (2003) demonstrated that the effects of models implemented in high schools and middle schools are about on par with those found in elementary schools, there are fewer research-based reform models that are available for high school educators to choose. It is also the case that the outcomes of interest for high school students are often quite different from those expected of elementary school students. For instance, staying in school and graduating and making a successful transition from high school to college or the workforce are critical outcomes that are unique to students attending school at the secondary level.

After years of largely being ignored, high school reform has recently gained greater attention in policy circles. A groundswell of reports has drawn attention to the problems of many American high schools, particularly those in large urban and high-poverty areas. For instance, Balfanz and Letgers (2004) reported that within nearly 1,000 high schools in the country, graduation is at best a 50/50 proposition. In 2,000 high schools, the freshman class shrinks by 40% or more by the time the students reach their senior year. According to Swanson (2004), nearly 1 out of 3 public high school students in the U.S. fails to graduate, and students from historically disadvantaged minority groups, including American Indian, Hispanic, and African American have little more than a 50% chance of completing high school and earning a diploma. Even for those minority students who do complete high school, Greene and Forster (2003) add that only 20% of

all African American students and 16% of Hispanic students leave high school ready for college.

A framework for understanding the reform of American high schools, which has been espoused by the Bill and Melinda Gates Foundation and also supported by the empirical work of MDRC, promotes schools founded on “three Rs:” “rigor;” “relevance,” and “relationships.” Rigor indicates that students have access to and take what is commonly known as a college preparatory curriculum. This tenet suggests that all students, regardless of their abilities or performance levels, should take four years of English and at least three years of mathematics, science, social studies, and a foreign language. If all students are to take these courses and pass them, it means that schools and the adults in them must commit to finding ways to help all students master these new basics—which usually means spending more time helping lower-performing students. “Rigor” means that all students will be prepared for postsecondary education without the need for remediation at the college level, and that there is an alignment between high school exit exams and postsecondary entrance requirements. It also means that expectations for all students are heightened and no students are relegated to low-level general or vocational track classes. It also means teachers must be fully qualified and competent in their discipline.

The idea of “relevance” shifts the focus to students and what motivates them to learn. Students in schools in which learning is relevant do not ask the question: “Why do I have to learn this?” Curricula is set in context so students can see how knowledge builds on what they already know, and it is applied so they can see how it is used in the real world. Studies are connected to students’ goals, and teachers and counselors help students

plan their course taking to meet their interests and career and college goals. Most importantly, students become engaged in their learning because they are able to see that what they are learning has meaning for them and will impact their futures. Finally, “relationships” speaks to breaking down the impersonalized nature of many large high schools to more personalized institutions in which students are more connected to their teachers and peers. The design of organizational structures, such as smaller learning communities or career academies, is meant to make students feel less anonymous and more engaged in their classes and the school.

One intervention with rigorous evidence of success that also fits this archetype of rigor, relevance, and relationships is the Career Academy concept. Typically serving 150-200 students in grades 9 or 10 through grade 12, Career Academies have three distinguishing features: (1) they are organized as small learning communities to create a supportive, personalized learning environment; (2) they combine academic and career and technical curricula around a career theme; and (3) they establish partnerships with local employers to provide career awareness and work-based learning opportunities for students. Operating as schools within schools and typically enrolling 30-60 students per grade, Career Academies are organized around themes including health, business and finance, and computer technology. Academy students take classes together, remain with the same group of teachers over time, follow a curriculum that includes both academic and career-oriented courses, and participate in work internships and other career-related experiences outside the classroom.

Over time, improving the rigor of academic and career-related curricula has become an increasingly prominent part of the Career Academies agenda. First established

more than 40 years ago, Career Academies operate today in more than 2,500 high schools across the country.

In urban high schools, too many students who manage to graduate are unprepared for postsecondary education or the world of work. These students, especially young men, often enter a labor market that offers them few opportunities for good jobs. Yet most high school reform efforts today focus solely on boosting academics. MDRC has conducted unusually rigorous evaluations of the effects of Career Academies on students' short- and long-term outcomes (Kemple & Willner, 2008). The findings related to this popular school reform initiative that combines academic offerings with career development opportunities shows that choosing between academics and career preparation is a false dichotomy. Career Academies produce sustained employment and earnings gains without sacrificing academics. In particular, Career Academies appear to offer young men a boost—comparable to the earnings premium of a year or two of postsecondary education—that puts them on a better earnings trajectory.

These results come from one of the first random assignment studies—the gold standard of program evaluation—ever conducted in a high school setting. MDRC has followed students in nine high schools around the country from when they entered 9th grade until eight years after their scheduled graduation. The results have particular relevance to historically underserved student populations, as more than 80 percent of students in the sample are African American or Hispanic. Most importantly, the MDRC researchers found that:

- Career Academies produced sustained earnings gains that averaged 11% (or \$2,088) more per year for program participants than for individuals in the control group—a \$16,704 boost in total earnings over the eight years of follow-up;

- These impacts on earnings are concentrated among young men and students at risk of academic failure. Young men saw an annual earnings gain of 17% (or \$3,731)—or nearly \$30,000 over eight years;
- This study shows that career development in high schools does not have to come at the expense of academic preparation. More than 90% of the students graduated from high school or received a General Educational Development (GED) certificate, and half earned a postsecondary degree or credential;
- Participants in Career Academies were more likely to be living independently with children and a spouse or a partner. Young men were more likely to be married (Kemple & Willner, 2008).

These empirical results, combined with the more theoretical foundation of rigor, relevance, and relationships, suggest that high schools that take the initiative to make schools and coursework more rigorous, more relevant to students' career goals, and more personalized and caring places—often through smaller career academies or theme-based schools within schools—enhance students' transition from high school to adulthood and the work force.

Beyond improving the rigor, relevance, and relationships within high schools, a variety of state and federal programs—such as Upward Bound and Talent Search—have focused on the transition from high school to college and, specifically, increasing college enrollment among historically underserved students. Given the sensitivity of college enrollment to tuition costs, it is logical to hypothesize that availability of financial aid programs should expand college-going. However, several studies examining the impact of the expansion of federal means-tested financial aid programs have found no evidence of increased college enrollment by low-income youth. These studies have analyzed college enrollment trends in the aftermath of the establishment of the Pell Grant program in the 1970s. Hansen (1983) first noted that there had been no disproportionate rise in college enrollment by low-income youth following implementation of Pell Grants. That

study has been criticized for relying too heavily on only two years of data and for including males whose decisions may have been affected by the end of the Vietnam War. However, later work by Kane and Avery (1994) confirmed that neither the choice of annual end-points nor the inclusion of males had significantly affected the findings. Manski (1993) also reported little evidence of a disproportionate growth in BA completion by low-income youth graduating from high school between 1972 and 1980.

One of the key explanations offered for this paradox is that lower-income students and their parents may not be fully informed of the cost of college and their eligibility for financial aid. Ikenberry and Hartle (1998), for example, found that public estimates of tuition costs were three times the actual costs. A second related explanation is that students from disadvantaged backgrounds have difficulty completing the sequential steps required to complete the college application process. Indeed, research by Kane and Avery (2004) demonstrated that low-income high school students have very little understanding of actual college tuition levels, financial aid opportunities, and how to navigate the admissions process. This lack of understanding of the costs of college and the procedures for application has been the target of several federally supported interventions.

In general, evaluations of programs like these have demonstrated the high cost of generating even modest increases in college-going among low-income youth. The Quantum Opportunity Program (QOP), an intensive demonstration program evaluated in the late 1990s, produced modest college enrollment effects by providing disadvantaged youth with a broad set of academic and support services throughout their high school years. QOP focused on providing targeted services, rather than formal instruction, to assist students in overcoming personal barriers to college attendance. A random

assignment evaluation of QOP showed that it did have a statistically significant impact on college enrollment, with program participants about three percentage points more likely to enroll in a two- or four-year college (Maxfield, Schirm, & Rodriguez-Planas, 2003). However, with per-student costs of \$4,000 to \$6,000 per year for four years, QOP's operational costs made widespread replication unlikely.

The Upward Bound program generated similarly modest impacts on college enrollment. Established in 1965 along with Title I and other War on Poverty programs, Upward Bound enrolls students in 9th grade who have low scholastic achievement and demonstrate a high likelihood for school dropout (Myers & Schirm, 1999). The program helps students prepare for and achieve success in postsecondary education through counseling, college application assistance, and supplemental academic instruction. After three years of follow-up, Myers et al. (2004) found the graduation rates and grade point averages were not impacted by assignment to the program for low-expectation students. Students defined as being at high academic risk received more high school credits in the Upward Bound program compared with high-risk students in the control group. This same gain was not observed for low-academic-risk students.

With respect to postsecondary education, the Upward Bound program had no impacts on enrollment rates of students or number of credits earned. However, for low-expectation students, those assigned to the Upward Bound treatment were more likely to attend a four-year college (38%) and to have earned more credits at a four-year college (21.9) compared with those in the control condition (18% rate of attendance and 11.0 credits). The Upward Bound program had especially strong impacts on the enrollment rates of Hispanics such that those in the treatment group were more likely to enroll in

postsecondary schools (50%) and earned more credits in postsecondary schools (28.4) than Hispanics in the control group (38% rate of attendance and 13.1 credits). Upward Bound was also found to impact student engagement in postsecondary schools. It increased the likelihood of student employment in college, the number of hours per week worked during college, receipt of personal counseling, attendance at learning skills centers, and use of tutoring services. Despite some evidence of effects, like the previously discussed QOP intervention, the Upward Bound program costs were relatively high at more than \$4,000 per student per year for up to five years.

A final promising program designed to help students make the transition from high school to college is the College Opportunity and Career Help (COACH) Program located at Harvard University's Kennedy School of Government (Kane, & Avery, 2004). The COACH Program was founded by Tom Kane and Christopher Avery, who were at the time both professors at the Kennedy School of Government at Harvard University. The program is in place at three high schools in Boston and provides inner-city students there with college admission, financial aid, and career path guidance. The coaches, who are graduate and undergraduate students at Harvard, work with small groups of students—most of whom are first-generation college students—helping them to navigate the obstacle course of the financial aid and college application process. The program is relatively inexpensive, with the student coaches being paid a stipend of \$1,200 under the expectation that they work three to four hours each week and participate in an intensive two-week training period.

Kane and Avery (2004) compared data on the participants of the COACH program with the data on one suburban, non-COACH school. They found that the inner-

city students and suburban students shared similar aspirations and perceptions of college costs and payoffs, although their participation rates in college differed significantly. Kane and Avery argue that the hurdles for inner-city youth often lie in the application process, such as registering for and taking the SAT, writing essays, and completing forms for financial aid. These findings further underline the conclusion that high school interventions that help students with their college applications may be more effective in improving college access for students from low-income families than increasing financial aid.

In addition to the successful transition from high school, staying in school and graduating from high school are certainly critical outcomes unique to secondary students. Recent research reviews conducted by the U.S. Department of Education's What Works Clearinghouse have provided some important evidence regarding the efficacy of replicable approaches to helping students stay in school, progress in school, and complete school (see <http://ies.ed.gov/ncee/wwc/reports/dropout/topic/>). As of September 2008, the What Works Clearinghouse had examined 84 studies of 22 dropout prevention interventions that qualified for review. Of these, 23 studies of 16 interventions met the rigorous research standards of the Clearinghouse—11 without any reservations and 12 of the studies with some reservations.

The review summarized outcomes across three domains for the 16 interventions: (a) staying in school, which included measures of whether the students stayed in school or dropped out without earning a diploma or GED; (b) progressing in school, which included measures of credits earned, grade promotion, and highest grade completed; and

(c) completing school, which included measures of whether a student earned a diploma or GED certificate.

In looking at the evidence compiled by the What Works Clearinghouse, three additional interventions had positive or potentially positive effects in two of the three domains considered:

- Accelerated Middle Schools had potentially positive effects on staying in school and positive effects on progressing in school;
- ALAS (Achievement for Latinos through Academic Success) had potentially positive effects on staying in school and on progressing in school;
- Check & Connect had potentially positive effects on staying in school and progressing in school.³

Eight other programs had potentially positive effects in one of three domains, and four had no discernable effects for any of the three domains.

Accelerated Middle Schools are self-contained academic programs designed to help middle school students who are behind grade level catch up with their age peers. If these students begin high school with other students their age, the hope is that they will be more likely to stay in school and graduate. The programs serve students who are one to two years behind grade level and give them the opportunity to cover an additional year of curriculum during their one to two years in the program. Accelerated Middle Schools

³ The What Works Clearinghouse also found that the Career Academies approach, which we discussed previously, had potentially positive effects on staying in school and on progressing in school. The report discussed earlier, which was authored by Kemple and Willner (2008), was produced very recently and was not included in the What Works Clearinghouse review, which was completed September 2008. Thus, my previous review of the Kemple and Willner report provides the most up-to-date information on the Career Academies model.

can be structured as separate schools or as schools within a traditional middle school. This model of reform is supported by three randomized controlled trials conducted by a research team from Mathematica Policy Research, who evaluated more than 800 students in school districts from Georgia, Michigan, and New Jersey (Dynarski, Gleason, Rangarajan, & Wood, 1998).

ALAS (“wings” in Spanish) is an acronym for Achievement for Latinos through Academic Success. ALAS is a middle school intervention designed to address student, school, family, and community factors that affect dropping out. The ALAS model calls for each student to be assigned a counselor who monitors attendance, behavior, and academic achievement. The counselor provides feedback and coordinates students, families, and teachers. Counselors also serve as advocates for students and intervene when problems are identified. Students are trained in problem-solving skills, and parents are trained in parent-child problem solving, how to participate in school activities, and how to contact teachers and school administrators to address issues. This intervention is supported by one study including 94 high-risk Latino students entering 7th grade in one urban junior high school in California (Larsen & Rumberger, 1995). The study, which measured student outcomes at the end of the intervention (9th grade) and two years after the intervention had ended (11th grade), revealed promising effects on whether students stayed in school and on the progress that they made in school.

Finally, the Check & Connect dropout prevention program relies on close monitoring of school performance, as well as mentoring, case management, and other supports. The Check component of the program is designed to assess student engagement through continuous and close monitoring of student performance and progress indicators.

The Connect component involves program staff giving individualized attention to students, in partnership with school personnel, family members, and community service providers. Students enrolled in Check & Connect are assigned a “monitor” who regularly reviews their performance—in particular, whether they are having attendance, behavior, or academic problems—and the monitor intervenes when problems are identified. The monitor also advocates for students, coordinates services, provides ongoing feedback and encouragement, and emphasizes the importance of staying in school. Two studies of Check & Connect included a total of more than 200 students attending Minneapolis high schools (Sinclair, Christenson, Evelo, & Hurley, 1998; Sinclair, Christenson, & Thurlow, 2005). In both studies the students entered the program at the beginning of the 9th grade. The researchers examined the program’s effects in the three dropout prevention domains considered by the What Works Clearinghouse, and it was found to have positive effects on staying in school and potentially positive effects on progressing in school.

Identifying Replicable Strategies with the Strongest Evidence of Effectiveness

When attempting to decide upon a single practice or program to implement in a school, or scale-up to serve multiple schools, one must weigh considerations regarding the costs, replicability, and quality of the evidence supporting the approach. Few research studies, or even whole bodies of evidence supporting a particular educational intervention, provide policymakers and practitioners with all three pieces of information, but this combination of evidence is essential for good decision making. An intervention backed by solid research demonstrating its effectiveness is worthless if it is too costly or too difficult to implement and scale up. Further, interventions that produce somewhat

slighter educational benefits can be preferable over others with evidence of greater benefits if the latter interventions are more expensive and more difficult to replicate. Reflecting on some examples from my work on CSR, I discuss how one may consider factors beyond a simple effect size when attempting to decide on the best available model for a given context.

Costs

Cost analysis and cost-effectiveness analysis in education help decisionmakers ascertain which program or combination of programs can achieve particular objectives at the lowest cost. As Levin (1995) noted, the underlying assumption is that different alternatives are associated with different costs and different educational results. By choosing those with the least cost for a given outcome, society can use its resources more effectively. By selecting more cost-effective approaches, those resources that are saved can be devoted to expanding programs. In this way, a systematic consideration of both costs and effects can help further the scale-up process.

In deciding whether or not to make a transition from a Title I schoolwide or targeted intervention model to an externally developed school reform model, a policymaker or practitioner may ask: are the benefits of implementing the model worth their seemingly high costs? On average, Borman et al. (2003) indicated that CSR programs have first-year costs of approximately \$85,000, including both personnel and non-personnel expenditures, which include items such as training and materials. However, some developers have argued that schools with concentrations of poor children generally are able to garner sufficient resources to implement CSR models by simply

reallocating existing supplemental funds and personnel from federal and state Title I programs, special education, desegregation settlements, and other sources (Slavin et al., 1994). In this way, many schools can afford even high-priced school reform models by simply trading in their largely remedial approaches of the past, most often represented by federal and state Title I programs, for new designs that will enable them to implement research-based schoolwide reform programs. As Odden & Archibald (2000) have argued, this method of “resource reallocation” can make implementations of programs essentially “costless.”

There are, indeed, clear challenges in determining the relative costs and benefits of CSR models (Levin, 2002), but if one assumes that implementations in high-poverty schools generally have few additional costs, the benefits suggested by the CSR meta-analysis are obviously well worth these modest investments. There is some research evidence to suggest that even if one does not assume that school reform implementations are “costless,” high-quality models are capable of yielding cost-benefit ratios that equal or exceed those found for other noted educational interventions, including the Tennessee Student/Teacher Achievement Ratio (STAR) class-size reduction effort (Borman & Hewes, 2003).

The analyses of Borman and Hewes revealed that a reform model that focuses on early intervention and prevention actually may save schools the investments in the costly remedial practices of special education referrals and retentions in grade, which can alone offset the costs of implementing the model. Though this evidence is important, much more cost-effectiveness research is needed for a wider range of school reform models, and for a broader array of educational interventions in general.

Replicability

Obviously, if one is concerned with implementing a promising program or practice in a school or scaling it up to serve a large number of schools, one must also consider the replicability of the programs and their effects. Borman and Hewes (2003), for instance, considered the replicability of four interventions with strong evidence of educationally meaningful impacts on students' short- and long-term outcomes: Success for All, the Perry Preschool, the Abecedarian Preschool, and the reductions in class size of the Tennessee Student/Teacher Achievement Ratio study.

Success for All and Perry Preschool are the two interventions of the four that are available as nationally disseminated models. Studies from diverse localities suggest that the educational effects of the original Success for All pilot programs tend to be replicated with a good deal of consistency, but that these effects depend on the quality of the implementation (Slavin & Madden, 2001). Implementation is not a trivial matter, as Success for All requires educators throughout a school to rethink and actively change many of their practices. After all, it is a whole-school reform model. If teachers do not accept the changes that the model suggests, it is not likely to succeed in improving practices and is not likely to affect student outcomes. Before adopting Success for All, the developer requires that 80% of the faculty agree, by secret ballot, to follow through with the implementation. If this support wanes, or if systemic support through the district or state tails off, the reform is likely to fail.

This has typically been the case in circumstances in which Success for All has failed, including the Memphis, Tennessee, school district, which dropped Success for All from more than 40 of its schools, and the Miami-Dade County school district, which

dropped the program from all but 7 of the 45 schools that once ran it. The overall quality of implementation, though, clearly is helped by the Success for All Foundation's national infrastructure for supporting schools that adopt the model and by federal policies, which make more supplemental resources available to finance school reform programs like Success for All.

Similarly, the educational approach used in the Perry Preschool classrooms and home visits is widely implemented today, primarily through the use of federal Head Start funds, as the High/Scope Curriculum (Epstein, 1993). Unfortunately, though, the significant resources necessary to replicate the Perry Preschool program, as it was originally designed in Ypsilanti, typically have not been available through publicly funded programs (Kagan, 1991; Barnett, 1995). There are other recent examples of high-cost, high-impact preschool programs, including the Chicago Child Parent Centers (Reynolds, Temple, Robertson, Mann, 2001), that have shown enduring effects on achievement and other important student outcomes. Examples such as these are significant in showing that the general concept of the intensive and relatively costly Perry Preschool model can be successfully funded and replicated. More public commitment through programs such as Head Start and Title I, or private support through community organizations and foundations, is needed to establish the large-scale national replication of the pilot program's effects.

Widespread efforts to deliver the Abecedarian model of highly intensive health, educational, and social services to children beginning shortly after birth have not been fully realized either. The Abecedarian project did inspire the U.S. Congress, in its reauthorization of the Head Start Act in 1994, to develop the Early Head Start program,

which covers the first three years of life. Since its inception, Early Head Start has grown to a nationwide effort of 635 community-based programs serving 45,000 children. However, similar to the comparison between Head Start and Perry Preschool, the Early Head Start program has not provided the same high-intensity services that the Abecedarian children received. Again, though the research evidence from the Abecedarian project clearly demonstrates that highly intensive early intervention can make a profound and enduring difference for the children who participate, the considerable monetary investments and capacity-building efforts to establish a similarly intensive national network of programs have not been undertaken by the federal government.

On the surface, the reductions in class size modeled by the Tennessee STAR study would seem to be the most easily replicated intervention of the four. During the Clinton administration, the federal government made available billions of dollars to reduce class sizes in the early grades. State-led efforts, such as California's massive initiative, also have provided support. At least two noteworthy differences, though, set apart the Tennessee STAR model from these national and state-level initiatives. First, the Tennessee STAR class-size reductions occurred in only those schools that had the facilities to accommodate the new classrooms needed to reduce class sizes. Second, the experiment operated in a relatively small number of schools and, therefore, did not create tremendous demands for new teachers.

As suggested by California's statewide initiative, scaling up class-size reductions to larger numbers of schools resulted in higher than anticipated costs, shortages of classroom space and qualified teachers, and smaller than anticipated achievement effects

(Bohrnstedt & Stecher, 1999). In addition, rather than improving equality of opportunity, Bohrnstedt and Stecher reported that the California effort exacerbated disparities between districts serving many minority and poor students and districts serving few minority and poor students. Therefore, in areas that require considerable capital improvements to make available the additional classroom space needed to reduce class sizes, and where there are potential shortages of qualified teachers, class-size reduction policies may not enjoy the level of success experienced in Tennessee.

Practical matters, including cost and the likelihood that an intervention's effects can be replicated and scaled up, should be considered along with careful analyses of the local context in which the program is to be implemented. Above, I mentioned some contextual factors that may hinder the replication of four model programs. These factors, along with cost information and general evidence of an intervention's replicability, should be considered by local policymakers when choosing among alternative approaches to improving the education of children from high-poverty contexts. For instance, local funding shortfalls would prevent faithful replication of the two preschool programs. Teacher shortages and a lack of additional classroom space might complicate class size reductions. Finally, a lack of commitment among teachers and principals to alter their practices and reform their schools may derail attempts to implement Success for All. All of these contextual issues, among many others that may be specific to the intervention or the locale in which it is to be implemented, may compromise the replication of promising interventions that could be put in place with relative consistency and reasonable monetary investment.

Evidence of Effectiveness

In identifying strategies for scale-up, one must simultaneously consider the overall quality, quantity, and effect size of the intervention. In reviewing the research base for replicable school reform programs, we developed appraisals of the evidence supporting 29 models for reforming high-poverty elementary and secondary schools. We defined four categories of the relative strength of evidence supporting each of the 29 models: *Strongest Evidence of Effectiveness*; *Highly Promising Evidence of Effectiveness*; *Promising Evidence of Effectiveness*; and *Greatest Need for Additional Research*.

With respect to the quality of the evidence, we sought to identify interventions that had the clearest causal relationships to student achievement outcomes. The level of confidence that the school reform model caused an improvement in student achievement depended on our ability to rule out other explanations for the increase in student achievement. We deemed, like Cook and Campbell (1979), the experimental and quasi-experimental research designs as among the most appropriate methodologies for ruling out alternative explanations. In addition to the suggestions of Cook and Campbell, we based this decision on our empirical results. That is, we found clear biases of one-group pretest-posttest designs relative to those studies that used experimental and quasi-experimental control groups.

The second key consideration when assessing the evidence base for an intervention, especially with regard to scale-up, is that there is a relatively large number of studies and observations from which one may generalize the findings for the

intervention to the population of schools in the U.S. that is likely to adopt and implement it. Establishing how many studies is enough to support claims that an educational program or practice is truly “scientifically based” is a bit more open to debate than decisions regarding the quality of the studies. In the instance of the meta-analysis of school reform effects, we used standards of 10 or more studies overall and 5 or more third-party control-group studies as the (arguably arbitrary) standards necessary to be in the top category.

Finally, in establishing the strength of an intervention’s evidence base, one must attempt to understand whether the outcomes are statistically significant, educationally meaningful, and, of course, positive. In the context of the meta-analysis of school reform effects, we asked: Does the evidence from control group studies show that the effects of the reform on student achievement are positive and statistically greater than 0? In establishing whether or not the effects were educationally meaningful, we compared the effects sizes for the school reform models to the effect sizes for various other existing standards and competing interventions. In the conclusion that follows, I return to this topic in attempting to understand the magnitude of the school reform effects by comparing them to different benchmarks.

The reform models meeting the highest standard of evidence—Direct Instruction, the School Development Program, and Success for All—are the only programs to have clearly established, across varying contexts and varying study designs, that their effects are relatively robust and that the models, in general, can be expected to improve students’ test scores. The models meeting the standard for the category of *strongest* evidence of effectiveness are distinguished from other available designs by the quantity and the

ability to generalize their outcomes, the quality of this evidence (for instance, six of the seven randomized experiments and many high-quality quasi-experimental control-group studies conducted on the models achieving the highest standard of evidence), and the reliable effects on student achievement. These programs are among the best examples of reforms being brought to scale that are likely to make a difference across large numbers of high-poverty schools.

Conclusion

In recent years, supplemental and whole-school reform models funded by Title I of the No Child Left Behind Act have been the primary federal policy initiatives at the forefront of the national movement to base the scale-up of educational reform on solid research evidence. This legislation, urging the use of research-based educational practices and procedures in schools receiving federal funding, has the potential to revolutionize school improvement in some of the most challenging contexts in the United States.

Does the quantity and quality of the research on whole-school reform and Title I provide the scientifically based evidence needed to identify the proven programs and practices that these new policies demand? What lessons might researchers, policymakers, and program developers learn from the preceding review of recent national efforts to bring reform to scale? Based on the prior review of findings from the two meta-analyses and the general history of national efforts to scale up reform in the nation's elementary and secondary schools, four clear implications emerge.

First, ironically, the two educational policy areas, CSRP and Title I, most recently and most strongly tied to higher standards of evidence, have clear limitations on the

overall quality and quantity of studies supporting their effects on achievement. Despite annual expenditures of approximately \$10 billion and a history of nearly 40 years, Title I itself has never been subjected to randomized trials (Borman & D'Agostino, 1996). Large-scale evaluations of Title I typically have provided nationally representative survey data describing the characteristics of Title I and non-Title I schools, the characteristics of Title I and non-Title I students, and the achievement outcomes of participants and non-participants. Quasi-experimental comparisons of the outcomes for Title I and non-Title I students have provided some insights into the potential achievement effects of the program. However, the results of these previous national evaluations ultimately suggest that researchers should focus less on attempting to generate national estimates of the program's characteristics and effectiveness and more on studying the effectiveness of specific interventions that could be funded under Title I.

Title I clearly is not a unique, supplemental, or uniform program. It is a funding mechanism designed to support a range of whole-school reform models, various instructional programs and practices, and school organizational and structural changes. Therefore much more may be learned by studying the effects of an array of replicable programs and practices. For example, in some states, it may be possible to permit a random sample of Title I schools to use their funds to reduce class sizes. Likewise, high-quality data on the effects of various whole-school reform models (e.g., Core Knowledge, Comer's School Development Program, or Success for All) could be generated by randomly selecting control and treatment sites from statewide lists of schools interested in implementing specific reform models. Another experimental strategy could involve multiple small-scale experiments, allowing for the investigation of multiple treatments.

The evidence provided by randomized field trials such as these could advance Title I research and policy in unprecedented ways.

The research on whole-school reform focuses on clear and replicable programs. The results, therefore, provide more direct implications for the scale-up of reform. The whole-school reform field, however, is still evolving. Twelve of 29 reform models are supported by five or more studies of their achievement effects, and only four models have been the subject of five or more third-party studies that used comparison groups. Over 40% of the analyses of CSR effects have been performed by the developers, and about half of the analyses have used some type of quasi-experimental control group. Only eight studies of four whole-school reform models have generated evidence from randomized experiments.⁴

Many of these problems are to be expected given the relatively recent emergence of whole-school reform models, in general, and many of the reform models, in particular. Some models are at an early stage of program development that has not yet demanded third-party evaluations and more costly and difficult control-group comparisons. On the other hand, there are some models that have had relatively long histories, have been replicated in many schools, and should have accumulated this evidence. Still other reform models are on their way to establishing a strong research base. Three models, in particular, have accumulated enough evidence to meet a relatively high standard of research evidence.

⁴ These reform models and studies include: the School Development Program (Cook, Habib, Phillips, Settersten, Shagle, & Degirmencioglu, 1999; Cook, Hunt, & Murphy, 1999); Direct Instruction (Crawford & Snyder, 2000; Grossen & Ewing, 1994; Ogletree, 1976; Richardson et al. 1978); Success for All (Borman et al., 2007); and Paideia (Tarkington, 1989).

Second, *this history of national efforts suggests a clear developmental trajectory from 1965 to the present that has resulted in historical improvement of disadvantaged students' outcomes. The implication seems to be that policy mandates and flexibility alone are less likely to produce educational reform and improved achievement outcomes than provider-based assistance to implement clear and replicable strategies for school change.* Though clearer federal mandates were associated with improved implementation and effects of Title I, these efforts were capable of producing no more than modest effects on student achievement outcomes. Despite continued efforts to tweak federal policy and provide greater flexibility to support school-based reform efforts, without a clear vision or model for reform, most schools did not capitalize on this flexibility.

In contrast, the most successful school reform models have enjoyed sustained periods of development, evaluation, and refinement and provide clear and replicable strategies for reforming schools. Despite being known as “comprehensive” models, the three most successful models focus on improvement in one rather discrete core area. Success for All and Direct Instruction have very clear instructional technologies that relate, most importantly, to improving literacy instruction. The School Development Program focuses its efforts on supporting students' holistic development to bring about academic success. In addition to a clear focus on improvement in a discrete area that the developer understands well, the models provide ongoing professional development and site-based assistance to help ensure the success of the reforms. These clear, focused, and well-supported school-based models of improvement are in stark contrast to top-down accountability mandates and flexibility for educational reform.

It is also the case that these externally developed, provider-assisted reforms contrast with traditional “home-grown” school reform models that have often characterized Title I and other efforts to reform elementary and secondary schools. The literature comparing these two types of reforms has rather consistently shown that the provider-assisted reforms tend to have stronger impacts than the home-grown models. This is supported by research dating back to the RAND Change Agent Study (Berman & McLaughlin, 1978; McLaughlin, 1990) to more recent work by Borman (2005) and others. According to the SEDL national CSR award database, the top 25-30 externally developed, provider-assisted reforms represent approximately 60% of all of the implementations of CSR. So schools do use a core group of reforms to implement and sustain the majority of reforms operating in U.S. schools. Also, these reforms have the benefit of being replicable and “scaleable” to serve many other U.S. schools, which makes evidence regarding their effects particularly relevant for policy.

Third, *the results from these national efforts suggest that large-scale reform is capable of producing widespread, but modest, achievement effects.* Historically, teaching has been fraught with what Lortie (1975) called “endemic uncertainties.” Moreover, Cook and Payne (2002) argued that the dominant perspectives on evaluation and improvement in education suggest that the context of each district, school, and classroom is so distinctive that only highly specific change strategies mapped to site-specific circumstances are likely to modify and improve their central functions. The continued growth of evidence-based policy, which has advanced the application of replicable technologies that are based on scientific knowledge, provides a clear contrast to these long-standing theories and beliefs about schools, educational change, and evaluation.

The successful expansion of CSR and other evidence-based practices shows that research-based models of educational improvement can be brought to scale across many schools and children from varying contexts. There are adaptations that are sensitive to context—for instance there is a Spanish version of the Success for All program, *Éxito Para Todos*, for English language learners—but the general models of school improvement also include well-founded and widely applicable instructional and organizational components that are capable of being brought to scale across a large number of schools. The previous growth in the market place of school reform models and the proven replicability of many of the programs are important developments. To further advance research-based practice, policymakers and educators must demand clear evidence that the reforms will make a difference.

The results from the meta-analyses suggest that the achievement effects associated with Title I and CSR are statistically significant, meaningful, and appear to have increased in magnitude as the policies and programs have been better implemented. Our various analyses suggest that Title I and CSR schools can be expected to score between nearly one-tenth and one-seventh of a standard deviation, or between 1.9 NCEs and 3.2 NCEs, higher than control schools on achievement tests. The low-end estimate represents the overall CSR effect size of $d = .09$ for third-party studies using comparison groups, and the high-end estimate represents the effect size of $d = .15$ for all evaluations of the achievement effects of CSR. Using U_3 , a metric devised by Cohen (1988), the effect size of $d = .12$ for all studies using control groups tells us that the average school implementing a CSR program outperformed about 55% of similar control schools that did not implement a CSR model.

How should we interpret this overall effect? Cooper (1981) has suggested a comprehensive approach to effect size interpretation that uses multiple criteria and benchmarks for understanding the magnitude of the effect. First, and most generally, we may compare the overall CSR effect size to Cohen's (1988) definitions of a small effect within the behavioral sciences, $d = .20$, and a large effect, $d = .80$. Second, and more specifically, Cohen (1988) pointed out that the relatively small effects of around $d = .20$ were most representative of fields that are closely aligned with education, such as personality, social, and clinical psychology. Similarly, Lipsey and Wilson's (1993) more recent compendium of meta-analyses concluded that psychological, educational, and behavioral treatment effects of modest values of even $d = .10$ to $d = .20$ should not be interpreted as trivial. Finally, and even more specifically, the effects of recent CSR models appear somewhat stronger than the effects of the extra resources and programs provided through Title I.

Fourth, high schools have historically been under funded by the federal government and have faced particular challenges for reform, yet those that have stressed the rigor and relevance of their course offerings, and those that have offered more personalized organizational structures that have enhanced the relationships within them appear to be directly linked to improved career and postsecondary outcomes. Further, various interventions have highlighted the importance of providing low-income and first-generation college students basic information concerning the costs of college and the procedures for application. These interventions, when carried out through cost-effective strategies, can show promise in helping students access postsecondary opportunities and in making successful transitions to college. Along with these specific programmatic

recommendations, I suggest that high schools also become more accountable and more data-based institutions. Prior federal mandates through NCLB have placed accountability front and center in grades 3 through 8, but beyond middle school there are few consistent indicators of students' and schools' progress. Greater accountability would help high schools monitor student progress and would place more reliable measures in the hands of policymakers so they can make difficult decisions regarding improving their course offerings and organizational effectiveness. If there are to be stronger linkages between America's schools and the workplace, then greater investments are needed within our country's high schools in terms of both their programmatic offerings and their accountability for student learning.

Finally, *better evidence is needed to provide both summative and formative appraisals of current and future national efforts to scale up reform in high-poverty elementary and secondary schools.* There are models that have been well researched and have shown that they are effective in improving student achievement across reasonably diverse contexts. These models certainly deserve continued dissemination and federal support through Title I and other federal programs. All school reform models—even those achieving the highest standard of evidence—would benefit from more federal support for the formative and summative evaluations that are necessary to establish even more definitively what works, where, when, and how.

Clear research requirements, ample funding for research and development, and a focus on the reform models' results may support the transformation of educational research and practices in much the same way that it has helped transform medical research and treatment. Like the series of studies required in the Food and Drug Administration's

premarketing drug approval process, a similar set of studies might guide the research, development, and ultimate dissemination of educational programs (Borman, 2003). Once a school reform program has met a standard of evidence, then its implementation using federal funds, most significantly those from Title I, should be approved. Before programs have accumulated such evidence, some concern should be shown for the ethics of supporting educational programs with unknown potentials. In medicine, Gilbert, McPeck, and Mosteller (1977) noted that only half of the new treatments subjected to randomized clinical trials actually showed benefits beyond the standard treatments patients would have received. Without the benefit of high-quality evaluation, many widely disseminated educational practices may simply waste the time of teachers and students or, potentially, do harm.

At the same time, schools and policymakers should not dismiss promising programs before knowing their potential effects. Instead, developers and the educational research community need to make a long-term commitment to research-proven educational reform and to establish a marketplace of scientifically based models capable of bringing widespread reform to the nation's schools. Similar to Donald Campbell's (1969) famous vision of the "experimenting society," we must take an experimental approach to educational reform, an approach in which we continue to evaluate new programs designed to address specific problems, in which we learn whether or not these programs make a difference, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available.

References

- Balfanz, R., & Letgers, N. (2004). *Locating the dropout crisis*. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.
- Barnett, W.S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5(3), 25-50.
- Berman, P., & McLaughlin, M.W. (1978). Federal programs supporting educational change, Vol. VIII: *Implementing and Sustaining Innovation*. Santa Monica, CA: RAND.
- Bohrnstedt, G.W., & Stecher, B.M. (Eds.) (1999). *Class-size reduction in California: Early evaluation findings, 1996-1998*. (CSR Research Consortium, Year 1 Evaluation Report). Palo Alto, CA: American Institutes for Research.
- Borman, G.D. (2005). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. In L. Parker (Ed.), *Review of Research in Education*, 29 (pp. 1-28). Washington, DC: American Educational Research Association.
- Borman, G.D. (2003). Experiments for educational evaluation and improvement. *Peabody Journal of Education*, 77(4), 7-27.
- Borman, G.D. (2000). Title I: The evolving research base. *Journal of Education for Students Placed At Risk*, 5, 27-45.
- Borman, G.D., & D'Agostino, J.V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 4, 309-326.
- Borman, G.D., & D'Agostino, J.V. (2001). Title I and student achievement: A quantitative synthesis. In G.D. Borman, S.C. Stringfield, & R.E. Slavin (Eds.), *Title I:*

- Compensatory education at the crossroads* (pp. 25-58). Mahwah, NJ: Lawrence Erlbaum Associates.
- Borman, G.D., D'Agostino, J.V., Wong, K.K., & Hedges, L.V. (1998). The longitudinal achievement of Chapter 1 students: Preliminary evidence from the Prospects study. *Journal of Education for Students Placed At Risk, 3*, 363-399.
- Borman, G.D., & Hewes, G.M. (2003). The long-term effects and cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis, 24*, 243-267.
- Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125-230.
- Borman, G.D., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal, 44*, 701-731.
- Bowler, M., & Thomas, D. (2006). *Regional educational laboratories awarded* (press release). Retrieved June 2, 2008 from <http://www.ed.gov/news/pressreleases/2006/03/03282006.html>
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409-429.
- Cohen, D.K. (1982). Policy and organization: The impact of state and federal education policy on school governance. *Harvard Educational Review, 52*, 474-499.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Comer, J.P. (1988). Educating poor minority children. *Scientific American, 259*(5), 42-48.

- Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cook, T.D., Habib, F.N., Phillips, M., Settersten, R.A., Shagle, S.C., & Degirmencioglu, S.M. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, 36, 543-597.
- Cook, T.D., & Payne, M.R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 150-178). Washington, DC: Brookings.
- Cooper, H. (1981). On the effects of significance and the significance of effects. *Journal of Personality and Social Psychology*, 41, 1013-1018.
- Crandall, D.P., Loucks-Horsley, S., Baucher, J.E., Schmidt, W.B., Eiseman, J.W., Cox, P.L., et al. (1982). *Peoples, policies, and practices: Examining the chain of school improvement* (Vols. 1-10). Andover, MA: The NETWORK.
- Crawford, D.B., & Snider, V.E. (2000). Effective mathematics instruction: The importance of curriculum. *Education and Treatment of Children*, 23(2), 122-142.
- Datnow, A., Borman, G., & Stringfield, S. (2000). School reform through a highly specified curriculum: A study of the implementation and effects of the Core Knowledge sequence. *The Elementary School Journal*, 101, 167-192.
- Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed At Risk*, 5, 183-204.

- Dynarski, M., Gleason, P., Rangarajan, A., & Wood, R. (1998). *Impacts of dropout prevention programs: Final report. A research report from the School Dropout Demonstration Assistance Program evaluation*. Princeton, NJ: Mathematica Policy Research, Inc.
- Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15-24.
- Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (1965).
- Epstein, A.S. (1993). *Training for quality: Improving early childhood programs through systematic inservice training*. Ypsilanti, MI: High/Scope Press.
- Finn, J.D. & Achilles, C.M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21, 97-109.
- Fullan, M.G. (with S. Stiegelbauer) (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Gilbert, J., McPeck, B., & Mosteller, F. (1977). Statistics and ethics in surgery and anesthesia. *Science*, 198, 684-689.
- Glass, G.V., & Smith, M.L. (1977). *"Pullout" in compensatory education*. Boulder, CO: University of Colorado, Laboratory of Educational Research.
- Greene, J., & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States*. New York: Manhattan Institute.
- Grissmer, D.W., Flanagan, A., & Williamson, S. (1998). Why did the black-white score gap narrow in the 1970s and 1980s? In C. Jencks, & M. Phillips (Eds.), *The black-white test score gap* (pp. 182–226). Washington, DC: Brookings Institution.

- Grossen, B., & Ewing, S. (1994). Raising mathematics problem-solving performance: Do the NCTM teaching standards help? *Effective School Practices*, 13(2), 79-91.
- Hansen, W. (1983). Impact of student aid on access. In J. Froomkin (ed.), *The crisis in higher education*. Washington, DC: Academy of Political Science.
- Hawkrige, D.G., Campeau, P.L., DeWitt, K.M., & Trickett, P.K. (1969). *A study of further selected exemplary programs for the education of disadvantaged children*. Palo Alto, CA: American Institutes for Research.
- Hawkrige, D.G., Chalupsky, A.B., & Roberts, A.O.H. (1968). *A study of selected exemplary programs for the education of disadvantaged children*. Palo Alto, CA: American Institutes for Research.
- Haynes, N., Emmons, C., & Woodruff, D. (1998). School Development Program effects: Linking implementation to outcomes. *Journal of Education for Students Placed At Risk*, 3, 71-86.
- Herrington, C.D., & Orland, M.E. (1992). Politics and federal aid to urban school systems: The case of Chapter 1. In J. Cibulka, R. Reed, & K. Wong (Eds.), *The politics of urban education in the United States* (pp. 167-179). Washington, DC: Falmer Press.
- Hirsch, E.D., Jr. (1995). *Core Knowledge Sequence*. Charlottesville, VA: Core Knowledge Foundation.
- Hirsch, E.D., Jr. (1996). *The schools we need*. New York: Doubleday.
- Ikenberry, S., & Hartle, T. (1998). *Too little knowledge is a dangerous thing: What the public thinks and knows about paying for college*. Washington, DC: American Council on Education.

- Jeffrey, J.R. (1978). *Education for children of the poor: A study of the origins and implementation of the Elementary and Secondary Education Act of 1965*. Columbus, OH: Ohio State University Press.
- Kagan, S.L. (1991). Excellence in early childhood education: Defining characteristics and next-decade strategies. In *The care and education of America's young children: Obstacles and opportunities. Ninetieth yearbook of the National Society for the Study of Education*, S.L. Kagan (Ed.), pp. 237-258. Chicago, IL: National Society for the Study of Education.
- Kane, T.J., & Avery, C. (2004). Student perceptions of college opportunities: The Boston COACH program. In Caroline Hoxby (ed.), *College decisions: The new economics of choosing, attending, and completing college*. Chicago: University of Chicago Press.
- Kearns, D. & Anderson, J. (1996). Sharing the vision: Creating new American schools. In S. Stringfield, S. Ross, & L. Smith (Eds.), *Bold plans for school restructuring* (pp. 9-23). Mahwah, NJ: Erlbaum.
- Kemple, J.J., & Willner, C.J. (2008). *Career academies long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York: MDRC.
- Kirst, M., & Jung, R. (1982). The utility of a longitudinal approach in assessing implementation: A thirteen-year review of Title I, ESEA. In W. Williams, R.F. Elmore, J.S. Hall, R. Jung, M. Kirst, S.A. MacManus, B.J. Narver, R.P. Nathan, & R.K. Yin (Eds.), *Studying implementation* (pp. 119-148). Chatham, NJ: Chatham House.
- Larson, K.A., & Rumberger, R.W. (1995). ALAS: Achievement for Latinos through academic success. In H. Thornton (Ed.), *Staying in school. A technical report of three*

dropout prevention projects for junior high school students with learning and emotional disabilities. Minneapolis, MN: University of Minnesota, Institute on Community Integration.

Letgers, N.E., Balfanz, R., Jordan, W.J., & McPartland, J.M. (2002). *Comprehensive reform for urban high schools: A talent development approach*. New York: Teachers College Press.

Levin, H.M. (1995). Cost-effectiveness analysis. In M. Carnoy (Ed.) *International encyclopedia of economics of education* (2nd ed.) (pp. 381-386). Oxford: Pergamon.

Levin, H.M. (2002). *The cost effectiveness of whole school reforms*. ERIC Clearinghouse on Urban Education, Urban Diversity Series 114. New York: Teachers College, Columbia University.

Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

Lortie, D.C. (1975). *Schoolteacher*. Chicago: University of Chicago Press.

Manski, C.F. (1992-93). Income and higher education. *Focus*, 14(3).

Maxfield, M., Schirm, A., & Rodriguez-Planas, N. (2003). *The Quantum Opportunity Program Demonstration: Implementation and short-term impacts*. Washington, DC: Mathematica Policy Research, Inc.

McLaughlin, M.W. (1976). Implementation of ESEA Title I: A problem of compliance. *Teachers College Record*, 77, 397-415.

McLaughlin, M.W. (1990). The Rand Change Agent Study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(9): 11-16.

- Meyer, J.W., Scott, W.R., & Strang, D. (1986). Centralization, fragmentation, and school district complexity. *Administrative Science Quarterly*, 32, 186-201.
- Murphy, J., & Beck, L. (1995). *School-based management as school reform: Taking stock*. Newbury Park, CA: Corwin.
- Myers, D., Olsen, R., Seftor, N., Young, J., & Tuttle, C. (2004). *The impacts of regular Upward Bound: Results from the third follow-up data collection*. Washington DC: U.S. Department of Education, Planning and Evaluation Services.
- Myers, D.E., & Schirm, A.L. (1999). *The impacts of Upward Bound: Final report for phase I of the national evaluation* (Executive Summary; Contract No.: LC-92001001; MPR Reference No. 8046-515). Washington, DC: U.S. Department of Education, Planning and Evaluation Services.
- No Child Left Behind Act of 2001, Pub L. No. 107-110, 115 Stat, 1425 (2002).
- Odden, A., & Archibald, S. (2000). *Reallocating resources: How to boost student achievement without asking for more*. Thousand Oaks, CA: Corwin.
- Ogletree, E. J. (1976). A comparative study of the effectiveness of DISTAR and eclectic reading methods for inner-city children. Chicago: Chicago State University. (ERIC Document Reproduction Service No. ED 146544).
- Olson, L., & Viadero, D. (2002). Law mandates scientific base for research. *Education Week*, 21(20), 1, 14-15.
- Peterson, P.E., Rabe, B.G., & Wong, K.W. (1986). *When federalism works*. Washington, DC: Brookings.
- Puma, M.J., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes*. Bethesda, MD: Abt Associates.

- Reynolds, A.J., Temple, J.A., Robertson, D.L., Mann, E.A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest—A 15-year follow-up of low-income children in public schools. *Journal of American Medical Association*, 285, 2339-2346.
- Richardson, E., Dibenedetto, B., Christ, A., Press, M., & Winsbert, B. (1978). An assessment of two methods for remediating reading deficiencies. *Reading Improvement*, 15(2), 82-95.
- Rotberg, I.C., Harvey, J., & Warner, K.E. (1993). *Federal policy options for improving the education of low-income students. Vol. I: Findings and recommendations*. Santa Monica, CA: RAND.
- Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. In C.B. Cazden (Ed.), *Review of research in education* (pp. 353-389). Washington, DC: American Educational Research Association.
- Sinclair, M.F., Christenson, S.L., Evelo, D.L., & Hurley, C.M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, 65 (1), 7–21.
- Sinclair, M.F., Christenson, S.L., & Thurlow, M.L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children*, 71 (4), 465–482.
- Sizer, T.R. (1992). *Horace's school: Redesigning the American high school*. New York: Houghton Mifflin.
- Slavin, R.E., & Madden, N.A. (2001). *One million children: Success for All*. Thousand Oaks, CA: Corwin.

Slavin, R.E., Madden, N.A., Dolan, L.J., Wasik, B.A., Ross, S.M., & Smith, L.M. (1994).

‘Whenever and wherever we choose’ The replication of ‘Success for All.’ *Phi Delta Kappan*, 75, 639-647.

Smith, M.S., & O’Day, J. (1991). Systemic school reform. In S.H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing, Politics of Education Association yearbook, 1990* (pp. 233-267). London: Taylor & Francis.

Stringfield, S., Millsap, M., Yoder, N., Schaffer, E., Nesselrodt, P., Gamse, B., et al. (1997). *Special strategies studies final report*. Washington, DC: U.S. Department of Education.

Swanson, C. (2004). *Projections of 2003-04 high school graduates*. Washington, DC: The Urban Institute.

Tarkington, S.A. (1989). *Improving critical thinking skills using Paideia seminars in a seventh-grade literature curriculum*. Unpublished doctoral dissertation, University of San Diego.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London: Falmer Press.

Wargo, M.J., Campeau, P.L., & Tallmadge, G.K. (1971). *Further examination of exemplary programs for educating disadvantaged children*. Palo Alto, CA: American Institutes for Research.

Whitehurst, G. (2002, April). *Charting a new course for the U.S. Office of Educational Research and Improvement*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

- Wong, K., & Meyer, S. (2001). Title I schoolwide programs as an alternative to categorical practices: An organizational analysis of surveys from the Prospects study. In G.D. Borman, S.C. Stringfield, & R.E. Slavin (Eds.), *Title I: Compensatory education at the crossroads* (pp. 195-234). Mahwah, NJ: Erlbaum.
- Wong, K.K., & Meyer, S.J. (1998). Title I schoolwide programs: A synthesis of findings from recent evaluation. *Educational Evaluation and Policy Analysis*, 20, 115-136.

