

Renewing our Schools,
Securing our Future



**Evidence-Based Reform:
Advancing the Education of Students at Risk**

Robert E. Slavin
Johns Hopkins University

Report prepared for:
Renewing Our Schools, Securing Our Future
A National Task Force on Public Education

A joint initiative of the Center for American Progress and the Institute for America's Future.

March, 2005



EXECUTIVE SUMMARY

Despite some recent improvements, the academic achievement of American students remains below that of those in most industrialized nations, and the gap between African American and Hispanic students and White students remains substantial. For many years, the main policy response has been to emphasize accountability, and No Child Left Behind has added further to this trend. There is much controversy about the effects of accountability systems, but they have had little impact on the core technology of teaching: Instruction, curriculum, and school organization.

This paper argues that genuine reform in American education depends on a movement toward evidence-based practice, using the findings of rigorous research to guide educational practices and policies. No Child Left Behind gives a rhetorical boost to this concept, exhorting educators to use programs and practices “based on scientifically-based research.” In practice, however, programs that particularly emphasize research-based practice, such as Reading First, have instead supported programs and practices (such as traditional basal reading textbooks) that have never been evaluated, while ignoring well-evaluated programs. The same is true of the earlier Comprehensive School Reform program, which was intended for “proven, comprehensive” programs but has instead primarily supported unresearched programs.

Despite these false starts, the evidence-based policy movement remains the best hope for genuine reform in U.S. education. The Institute for Education Sciences in the U.S. Department of Education, as well as NICHD, NSF, and other funders, are supporting many research and development initiatives that use rigorous randomized experiments to evaluate educational products and practices. Of equal importance, the What Works Clearinghouse (WWC) is beginning to review educational programs to identify those supported by rigorous research. These changes create the possibility that educators will soon have available a broad range of programs in which they can place confidence, just as Food and Drug Administration approval gives physicians and the public confidence in medical treatments.

This paper reviews research on programs that already have strong evidence of effectiveness. It establishes criteria for study quality like those of the WWC. Programs with strong evidence of effectiveness fell into the following categories.

1. Comprehensive school reform models, which provide professional development and materials to improve entire schools. Research particularly supports *Success for All* and *Direct Instruction*, but smaller numbers of studies support several additional models including the *School Development Program*, *America’s Choice* and *Modern Red Schoolhouse*.
2. Instructional technology. Research supports integrated learning systems in mathematics. Word processing has been found to improve writing achievement.
3. Cooperative learning programs engage students in small groups to help each other learn. Many studies support this strategy in elementary and secondary math, reading, and other subjects.

4. Innovative mathematics programs. The first What Works Clearinghouse report supported research on two technology-based programs, *Cognitive Tutor* and *I Can Learn*, in middle schools. Elementary programs such as *Cognitively Guided Instruction* and *Project SEED* also have strong evidence of effectiveness.
5. Innovative elementary reading programs having strong evidence of effectiveness include *Success for All* and *Direct Instruction*, as well as *Reciprocal Teaching* and *Cooperative Integrated Reading and Composition*.
6. Tutoring programs in reading, especially *Reading Recovery*, have rigorous evaluations showing their effectiveness.
7. Dropout prevention programs, such as the *Coca-Cola Valued Youth Program* and *Alas*, have good evidence of effectiveness.

POLICY RECOMMENDATIONS

Clearly, much more research is needed, and current policies are not supporting use of the research-based programs that now exist. Policy recommendations are as follows.

1. Substantially increase support for research and development to at least \$500 million per year.
2. Fund development of new programs.
3. Fund evaluation of existing and new programs.
4. Provide incentives for schools to participate in research.
5. Provide incentives for schools and districts to use programs validated in rigorous research.
6. Maintain the integrity of proven programs by ensuring that publishers continue to provide professional development and support like that provided in the research.
7. Encourage states to base policies on research.

CONCLUSION

The solutions to America's education problems must draw on our nation's ingenuity, inventiveness, and technology. We can solve these problems as we have solved many others, by using research, development, and dissemination of effective tools and practices.

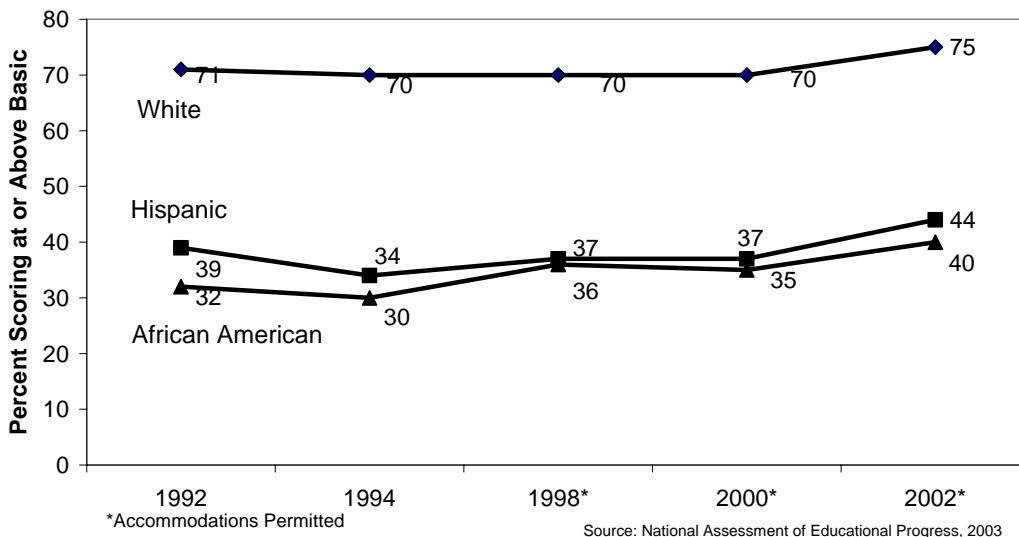
EVIDENCE-BASED REFORM

In 1983, *A Nation at Risk* declared that American schools faced a “rising tide of mediocrity,” and that America was in danger of falling behind its international competitors because of the poor performance of its students (National Commission on Excellence in Education, 1983). Since that time, American schools have been continuously engaged in reforms, mostly directed at increasing accountability among educators for student performance on state tests. Yet more than 20 years after *A Nation at Risk*, the achievement of U.S. students is virtually identical to what it was in the early 1980s. The National Assessment of Educational Progress (NAEP; Grigg, Daane, Jin, & Campbell, 2003) has shown small improvements in mathematics since 1980, but in reading the overall trend is virtually flat.

On international comparisons, U.S. students continue to score below most other industrialized nations, as shown most recently in the Programme for International Student Assessment (PISA), which placed U.S. 15-year-olds below those in most Asian and European countries on tests of mathematics, reading, and science (OECD, 2004). American students scored far below such similar nations as Canada, Australia, and New Zealand (Finland scored highest in most categories).

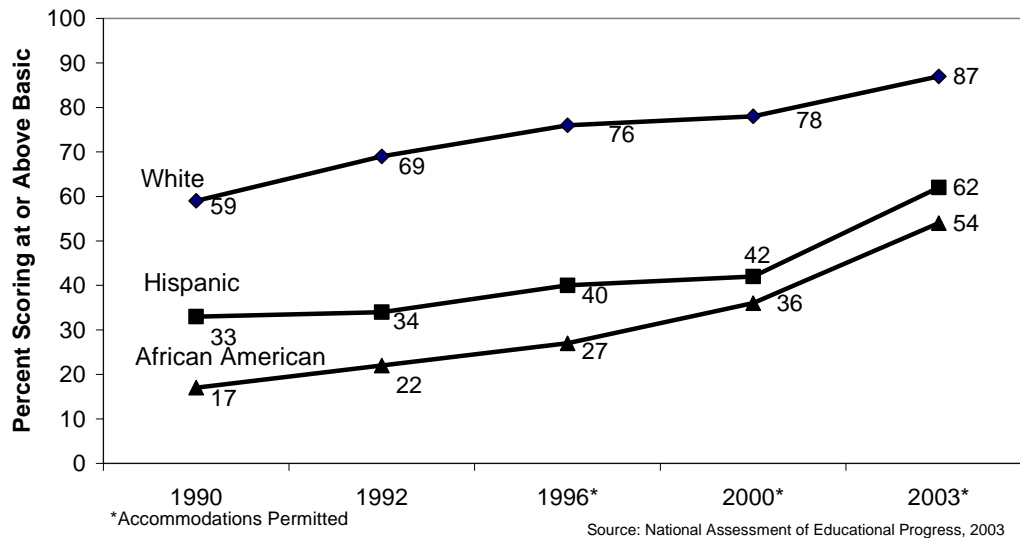
While overall achievement levels are a major concern, far more serious is the continuing gap between White children and African American, Hispanic, and Native American children. The gap in test performance on NAEP reading between African American and White fourth graders narrowed significantly during the 1970s, due primarily to improvements in the South. Yet since 1980, the gap has been virtually unchanged. Figure 1 shows trends in reading performance from 1992 to 2002 that illustrate how much of a gap remains and how little it has changed (data are from Grigg, Daane, Jin, & Campbell, 2003).

Figure 1
National Assessment of Educational Progress
Reading, 1992-2002



NAEP trends in mathematics, summarized in Figure 2, contrast with those in reading. Since 1990, mathematics performance for third and fifth graders has increased significantly (although U.S. students remain below other countries on international comparisons, such as TIMMS and PISA). However, the gap between White, Hispanic, and African-American students has persisted even as scores have risen.

Figure 2
National Assessment of Educational Progress
Mathematics, 1990-2003



The current national effort for reforming America’s schools is No Child Left Behind (NCLB), introduced by President Bush but supported by a bipartisan coalition. No Child Left Behind has many aspects, but its main focus is once again on accountability. It provides a variety of sanctions for schools that fail to meet “adequate yearly progress” goals on their state assessments. It is too soon, however, to know what effect NCLB will have on student achievement (see Center on Education Policy, 2003).

Beyond Accountability

There is considerable argument among researchers about the impacts of accountability on student learning. For example, Carnoy & Loeb (2002) found only slight differences on NAEP gains favoring states with strong accountability systems, while Neill & Gaylor (2001) and Amrien & Berliner (2003) found that states with strong accountability systems had *lower* gains on NAEP than other states. Regardless of this controversy, accountability is here to stay, and some level of accountability is likely to be a part of any rational policy to improve educational outcomes for at-risk children. However, reforms focusing on accountability and other management strategies have an inherent limitation. They do little to change the core technology of teaching, the materials, methods, and capabilities of front-line educators. In order to accelerate the improvement of educational outcomes for at-risk students, it is important to improve the quality of teachers, by improving salaries, working conditions, and teacher preparation (see Darling-Hammond, 1995), and to introduce innovations with strong evidence of effectiveness that do affect the core of instruction, curriculum, assessment, and school organization, where

education takes place. There is much research that suggests that minority students are particularly sensitive to the quality of instruction they receive (see Slavin & Madden, 2002). More than anyone else, minority students (and probably disadvantaged white students as well) need better teachers using better methods and materials every day.

Evidence-Based Practice

One aspect of No Child Left Behind touched on a movement in education reform that could finally move educational practice forward, and produce genuine and lasting improvements in student achievement. This is the movement toward evidence-based practice, use of the findings of rigorous research to guide educational practices and policies. In the No Child Left Behind legislation, educators were exhorted to use programs and practices “based on scientifically-based research” more than 100 times. This concept was particularly central to the Reading First program, which is providing almost \$1 billion per year to help low-performing schools improve reading instruction in Grades K-3. Previously, the 1997 Obey-Porter Act established a substantial fund to support schools in adopting “proven, comprehensive” reform models, and much other recent legislation has supported the concept that federal dollars should focus on programs with strong evidence of effectiveness, usually defined as ones that have been evaluated in randomized or matched experiments (see, for example, Whitehurst, 2002; Slavin, 2003).

In practice, the federal commitment to research-based practice has been more rhetorical than real. Few of the programs supported by Obey-Porter Comprehensive School Reform funding, for example, had any evidence of effectiveness. In fact, about half of the funding each year has supported various “home-grown” models assembled for the purpose. Reading First used the formulation “based on” scientifically-based research, and the states were given freedom to interpret this broadly. Universally, states have qualified traditional basal textbooks as “based on scientifically based research,” and as a result, the great majority of Reading First schools are using basal textbook programs that have never been evaluated in any sort of experiment.

Despite these false starts, the evidence-based policy movement remains the best hope for genuine reform of education in the U.S. Perhaps the most important reason for optimism in this regard is a set of extraordinary developments in the U.S. Department of Education’s Institute of Education Sciences (IES). Under the leadership of Grover Whitehurst, IES has radically changed its funding priorities to focus on the development and rigorous evaluation of practical, replicable programs designed to help educators achieve better outcomes with students. “Rigorous evaluation” primarily means studies in which students or schools are assigned at random to experience a given treatment or to serve as a control group. Random assignment experiments of this kind are the “gold standard” of research. They have long been the norm in medicine and other fields, but have been rare in education (see Mosteller & Boruch, 2002). By emphasizing randomized experiments and other rigorous experimental methods, Whitehurst and like-minded colleagues hope to erase the “awful reputation” of educational research (Kaestle, 1993) and to make a better case to Congress and other policymakers for increasing funding and support for far more research in education, and for using the findings of research as a basis for policy and practice.

At present, IES has funding programs under way to promote randomized experiments in a broad array of areas: early childhood, beginning reading, programs for struggling readers, secondary reading, math, science, education for English language learners, teacher professional development, after school programs, and more. The National Institute of Child Health and Human Development (NICHD) and the National Science Foundation (NSF) are also supporting such studies. Within a few years programs of all kinds will be emerging from this process with the kind of evidence of effectiveness that can serve as a solid base for policy and practice.

Alongside the move toward funding of R&D, IES is also sponsoring an effort that could ultimately become equally important. This is the What Works Clearinghouse (WWC), which is reviewing research on programs in a variety of areas to identify those that have strong evidence of effectiveness. The WWC has only recently issued its first report, on middle school mathematics, but the potential impact of this effort is considerable. Using consistent, widely supported standards for evaluating and synthesizing individual studies and bodies of research, the WWC can give educators and policymakers confidence that given programs or practices really are more effective than others.

The combined impact of the recent focus on rigorous evaluations and the What Works Clearinghouse could ultimately be revolutionary. Educational practice and policy has long ignored research for two main reasons. First, there has been too little high-quality research to serve as a basis for action. For example, a key reason that Reading First could not specify that schools must use programs that had been rigorously evaluated many times is that this would have in effect forced them to choose between only two programs with extensive evidence of effectiveness: Direct Instruction (Adams & Engelmann, 1996) and Success for All (Slavin & Madden, 2001). Having many programs with strong evidence would give policymakers the opportunity to recommend a list of programs, not just two.

Second, research has been ignored by educators on the basis that it is hard to know whose research is credible. Every salesman for any publisher or technology company claims to have evidence to support the company's products, and usually has a few charts showing some schools that made wonderful gains using their product. Educators know to take these charts with a grain of salt, but they do not feel capable of looking at competing claims and deciding which are justified. The new government-funded research, and especially the WWC, will soon begin to give an essential imprimatur to programs that have scientifically respectable evidence of effectiveness. Not every educator or policymaker has the time or skills to look into the literature on various programs, but anyone can understand a list of well-evaluated programs published by a respected, independent agency.

If it takes hold, the What Works Clearinghouse, and federal policies favoring use of programs validated by the WWC, will force publishers and technology companies to carry out their own research on their products, or to commission independent organizations to do so. This is exactly what happened with the establishment of the Food and Drug Administration, which primarily reviews evaluations carried out by drug companies. In combination with government-funded studies, new and increasingly effective methods and materials are likely to be created and then disseminated broadly. Even though federal funding remains a small part of national

education expenditures (about 7 percent), federal policies, information, and research can have influence far beyond this proportion.

In education, accountability for outcomes is necessary but not sufficient. Eventually, educators must also become accountable for using validated practices. In a recent article in *The New Yorker*, physician Atul Gawande (2004) described how scientific advances had raised life expectancies for children born with cystic fibrosis from three years in the 1950s to 40 years now, yet success rates today were found to vary substantially from hospital to hospital. The Cystic Fibrosis Foundation collects these data, and is beginning to make them available to physicians so they can benchmark their success rates against those of other hospitals, so that all hospitals can seek to emulate the best ones. Genuine, lasting progress will come in education (as in medicine) when practitioners have available effective methods, are expected to use them with intelligence and skill, and use data to monitor outcomes and benchmark these outcomes against those of practitioners in similar circumstances.

RESEARCH ON EFFECTIVE SCHOOL REFORM STRATEGIES

A strategy of using the findings of rigorous research as a basis for policy and practice depends on the existence of a substantial body of research that identifies practical, replicable models for school and classroom reform. Ideally, educators and policymakers should be able to choose among an array of solutions, each of which is known to be effective. These solutions should include programs for each subject and grade level, schoolwide issues such as assessment, classroom management, and attendance, subgroup issues such as accommodations for English language learners and students in special education, and whole-school comprehensive reform models. Unfortunately, research is thin in many of these areas, but in others there is a robust research base that can serve as a basis for evidence-based policy and practice. The following sections of this paper discuss the state of the evidence in several areas of educational practice that have been studied using rigorous experimental methods.

Criteria for Study Inclusion

This paper applies a consistent definition of “rigorous research” that is then used as a criterion for inclusion in the reviews. This definition is similar to the definitions used by the What Works Clearinghouse to define its top two categories of research quality.

1. The study had to compare an experimental treatment to a control group that received an alternate treatment or a treatment that represents common practice (which the experimental group would have received in the absence of the treatment).
2. The experimental and control treatments had to be equivalent before the treatments were applied. Ideally, students, classes, or schools were assigned at random to experimental or control conditions, and analysis was done at the level of random assignment. Studies of this kind would meet the highest study quality criterion set by the What Works Clearinghouse. Alternatively, the experimental and control groups could be matched based on variables such as prior achievement, socioeconomic status, ethnicity, and location. Such studies were included if pretest differences were less than 0.5 standard deviation units and if analysis of covariance, multiple regression, or similar procedures

were used to control for any pre-existing differences. Well-matched studies would generally meet the second study quality criterion set by the What Works Clearinghouse. In the present paper, randomized studies are emphasized when they exist, but since randomized evaluations are rare, well-matched studies are also considered as valid evidence of program outcomes.

3. Study duration had to be at least 10 weeks, preferably at least a year. Shorter studies may be of value in theory building, but evidence over a significant time period is essential for educational practice and policy, as brief studies often create artificial conditions that could not be maintained as long as a full academic year.
4. Quantitative measures of academic achievement had to be used.
5. Programs had to be evaluated with disadvantaged and minority students.

Effect Sizes

Whenever possible, outcomes of individual studies or groups of studies are characterized in terms of “effect sizes.” These are generally computed as the difference between the means of the experimental and control groups divided by the control group’s standard deviation. An effect size of +0.20 is considered a minimum for educational significance. At the high end, one useful benchmark is to note that studies of one-to-one tutoring in first grade reading by certified teachers averaged an effect size of +0.75 (Wasik & Slavin, 1993).

If studies did not provide sufficient data for computation of effect size, this was not a reason for exclusion. Instead, outcomes were characterized in other ways, but not averaged. In many cases, this paper reviews meta-analyses, which are quantitative syntheses of research in which effect sizes are averaged. Inclusion criteria and methods for computing effect sizes are generally similar to those described above, unless otherwise noted.

COMPREHENSIVE SCHOOL REFORM PROGRAMS

Comprehensive school reform models are methods designed to reform the most important aspects of school functioning: curriculum, instruction, programs for struggling students, assessment, school organization, parent involvement, professional development, and more. Several education reform organizations have developed, evaluated, and disseminated comprehensive school reform programs of many kinds, and this set of approaches has received substantial research attention over the past 15 years.

A review of experimental research on comprehensive school reform models was published by Borman, Hewes, Overman, & Brown (2003). A simplified adaptation of their main results appears in Table 1.

Table 1
Summary of Research on Comprehensive School Reform Models

	Number of Studies	(Third Party)
<i>Strongest Evidence of Effectiveness</i>		
Success for All	41	(25)
Direct Instruction	40	(38)
School Development Program	9	(5)
<i>Highly Promising Evidence of Effectiveness</i>		
Roots & Wings	5	(4)
Expeditionary Learning/Outward Bound	4	(3)
Modern Red Schoolhouse	4	(3)
<i>Promising Evidence of Effectiveness</i>		
Accelerated Schools	3	(2)
America's Choice	1	(1)
ATLAS Communities	2	(2)
Montessori	2	(2)
Paideia	3	(3)
The Learning Network	1	(1)
<i>Greatest Need for Additional Research</i>		
Audrey Cohen	1	(1)
Center for Effective Schools	0	(0)
Child Development Project	2	(0)
Coalition for Essential Schools	1	(1)
Community for Learning	0	(0)
Community Learning Centers	1	(1)
Co-Nect	5	(4)
Core Knowledge	6	(6)
Different Ways of Knowing	1	(1)
Edison	3	(3)
High Schools That Work	4	(0)
High/Scope	3	(2)
Integrated Thematic Instruction	1	(1)
MicroSociety	1	(0)
Onward to Excellence II	0	(0)
Talent Development High School	1	(0)
Urban Learning Centers	0	(0)
All CSR Models	145	(109)

Adapted from Borman, Hewes, Overman, & Brown (2003)

Borman et al. (2003) categorized programs according to the numbers of well-designed experiments on each and the consistency of positive achievement effects. The following sections discuss some of the most widely used and extensively researched of these models.

Success for All

Success for All (Slavin & Madden, 2001) is the most widely used and extensively evaluated of the CSR models. It provides schools with specific curriculum materials and extensive professional development in reading, writing, and language arts, along with detailed assessment, cross-grade grouping strategies, within-school facilitators, and other school organization elements. The program gives one-to-one tutoring to primary-grades children who are struggling in reading, and extensive outreach to parents. It provides detailed teacher's manuals and about 26 person-days of on-site professional development to enable schools to engage in a substantial retooling process. Originally focused on elementary school, prekindergarten to grade 6, Success for All now has a middle school (6-8) program as well. Programs in mathematics, science, and social studies were also developed, and the term Roots & Wings was used to describe schools using all of these elements (Slavin, Madden, Dolan, & Wasik, 1994). However, most schools, including many of those categorized as "Roots & Wings" in the Borman et al. (2003) review, use only the reading program, and the Roots & Wings term is no longer used. Research on Success for All and Roots & Wings are combined for discussion in this paper.

Borman et al. (2003) identified a total of 46 experimental-control comparisons evaluating Success for All, of which 31 were carried out by third-party investigators. A mean effect size of +0.20 (combining Success for All and Roots & Wings) was obtained across all studies and measures. A longitudinal study by Borman & Hewes (2003) found that students who had been in Success for All elementary schools were, by eighth grade, still reading significantly better than former control group students and were about half as likely to have been retained or assigned to special education.

Since the Borman et al. review, a number of additional studies of Success for All have been carried out. Most importantly, a national randomized evaluation of Success for All is under way. A total of 41 schools were randomly assigned to use Success for All either in grades K-2 or in grades 3-5. The primary grades in 3-5 schools were used as controls, as were the intermediate grades in K-2 schools. First-year results found positive effects for students in kindergarten and first grades on reading measures (Borman, Slavin, Cheung, Chamberlain, Madden, & Chambers, in press). Preliminary analyses of second-year results are finding stronger impacts. This first large-scale randomized evaluation is particularly important in today's policy environment, which is strongly supporting randomized experiments (Whitehurst, 2002). Taken together, there are now more than 50 experimental-control studies of Success for All involving more than 200 schools throughout the U.S.

Since 1998, Success for All has been developed and disseminated by the non-profit Success for All Foundation, and is currently working in about 1,400 schools in 47 states.

Direct Instruction

Direct Instruction (DI; Adams & Engelmann, 1996), once known as DISTAR, is an elementary school program originally designed to extend an effective early childhood curriculum into the early elementary grades, in a federal program called Follow Through. Like Success for All, DI is primarily intended to help high-poverty schools succeed with all students, and the program is even more systematically specified for teachers.

The DI reading and math programs have long been marketed by SRA, a division of the McGraw-Hill publishing company, under the titles “Reading Mastery” and “Connecting Math Concepts.” The publisher provides limited professional development with the program, but schools can contract with providers of professional development, primarily the National Institute for Direct Instruction (NIFDI) at the University of Oregon. Such schools receive approximately 32 person-days of professional development in their first year, similar to the services provided in the Follow Through studies. Research on DI has overwhelmingly focused on the model with extensive professional development, not on use of the books alone, and research findings for DI should therefore be assumed to apply only to the program with professional development. Certainly only this form could be considered a comprehensive reform model.

Borman et al. (2003) identified 40 experimental-control studies of DI, of which 38 were third party. The mean effect size was +0.15.

Other reviews of research on comprehensive school reform models have also identified Success for All and Direct Instruction as the CSR models that are most thoroughly supported by research. They include an influential review by the *American Institutes for Research* (Herman, 1999) and a review by *New York Times* reporter James Traub (1999).

School Development Program

James Comer, a Yale psychiatrist, developed one of the earliest of the comprehensive reform models, the School Development Program (SDP; Comer, Haynes, Joyner, & Ben-Avie, 1996). The focus of SDP is on the whole child. Rather than focusing on specified curricula and instructional methods, SDP concentrates on building a sense of common purpose among school staff, parents, and community, working through a set of teams in each school that develop, carry out, and monitor reforms tailored to the needs of each school. A School Planning and Management team develops an overall plan, and Mental Health and Parent teams focus on issues beyond the classroom.

Borman et al. (2003) listed SDP as one of three CSR programs with “strongest evidence of effectiveness,” even though the nine experimental-control studies that met their criteria produced a mean effect size of only +0.05. However, a set of three remarkable third-party evaluations provide better evidence of the program’s impact. One, a randomized evaluation in Prince George’s County, MD, found poor implementation and no achievement effects (Cook et al., 1999), but a partially randomized study in Chicago (Cook, Murphy, & Hunt, 2000) and a matched study in Detroit (Millsap, Chase, Obeidallah, Perez-Smith, & Brigham, 2000) found small but positive impacts on achievement.

Modern Red Schoolhouse

Modern Red Schoolhouse (Heady & Kilgore, 1996) is a program that emphasizes standards-based teaching, appropriate uses of technology, and frequent assessment. It provides customized professional development to help schools build coherent curricula aligned with state standards and then implement aligned practices. In recent years, Modern Red Schoolhouse has begun to focus more on district reform and leadership.

Borman et al. (2003) identified four experimental-control studies of Modern Red Schoolhouse, with an average effect size of +0.17.

Accelerated Schools

Accelerated Schools (Levin, 1987) is a process-oriented school reform model that emphasizes high expectations for children and giving students complex and engaging instruction. Each school staff designs its own means of putting into practice the basic principles.

Borman et al. (2003) identified three studies of Accelerated Schools with a mean effect size of +0.21.

Expeditionary Learning/Outward Bound

Expeditionary Learning (Campbell et al., 1996) is a design built around “learning expeditions,” which are “explorations within and beyond school walls.” The program is affiliated with Outward Bound and incorporates its principles of active learning, challenge, and teamwork. It makes extensive use of project-based learning, cooperative learning, and performance assessments.

Borman et al. (2003) identified four experimental-control evaluations of Expeditionary Learning, which had substantial positive effects.

America’s Choice

America’s Choice (NCEE, 2003) is a comprehensive reform model that focuses on standards and assessments, instruction aligned with standards, extensive professional development, and parent involvement. In particular, the program mandates a core curriculum in literacy and mathematics, tutoring for struggling students, and a school leadership team to coordinate implementation.

Borman et al. (2003) identified only one study of the America’s Choice design, but more recently researchers at the Center for Policy Research in Education at the University of Pennsylvania have carried out several evaluations. A longitudinal matched study in Rochester, NY, found that America’s Choice students made greater gains than other students from 1998 to 2003 in reading and math (May, Supovitz, & Perda, 2004). A matched study in Duval Co., Florida (Supovitz, Taylor, & May, 2002) compared America’s Choice and other schools on state tests, and results favored the AC schools in writing and, to a small degree, in math (but not reading). A one-year matched study (Supovitz, Poglinco, & Snyder, 2001) also compared matched AC and control schools in Plainfield, NJ, and found greater gains for the AC students on the state English Language Arts test.

INSTRUCTIONAL TECHNOLOGY

The use of various kinds of instructional technology to improve student achievement has become virtually universal in American education. In 1999, there were 8.6 million computers in U.S. schools and the ratio of students to computers was 5.7 to 1, up from 14 to 1 as recently as 1994 (U.S. Department of Education, 2002; Becker, 2001). Yet computers and other technology are used in many quite different ways. It is no longer meaningful to ask what the effects of

computer use on student achievement are instead, we must ask about the effects of each of several types of uses of technology (see Bebell et al., 2004; Blok et al., 2002).

Kulik (2003) recently reviewed research on learning effects of instructional technology in elementary and secondary schools. He carried out a meta-analysis, averaging effect sizes for studies comparing students using various technologies to those using non-technology methods to learn the same content. This paper reports his conclusions with adjustments to remove studies of very brief durations, and cites other recent reviews of particular applications of technology.

Integrated Learning Systems

Integrated learning systems are computer software systems that provide students with individualized courseware and management systems to present students with material that is at their appropriate level, keep records of their progress, and provide students with feedback and recognition for their learning efforts.

Kulik's review came to sharply different conclusions depending on the subject involved. Across 16 studies, ILS used in mathematics had a median effect size of +0.38. In contrast, nine studies of ILS used in reading found trivial effects, a median of +0.06. Most of the studies on both subjects involved use of Jostens (now called Compass Learning) ILS programs, and these had an effect size of +0.37 in math (nine studies) and +0.22 in reading (five studies). ILS programs appear to be more effective if students spend more time on them and if they work on them in well-structured pairs (see Lou, Abrami, & d'Appolonia, 2001).

Word Processing

By far the most common use of technology in education is word processing (Becker, 2001). Of course, the word processor has replaced the typewriter in schools and elsewhere, but the instructionally interesting question is whether use of word processors increases students' ability to write well, in comparison to learning to write using paper and pencil alone.

Reviews by Bangert-Drowns (1993) and Cochrane-Smith (1991) found positive effects of the use of word processors on writing skills. Kulik's (2003) review added four more studies, three of which were small and brief.

Other Applications of Technology

Kulik (2003) reviewed the achievement effects of several additional types of technology. Based on 12 studies of the Writing to Read computer program, he claimed this program had positive effects, although most studies were very small (30 to 97 students). Larger studies (121 to 1,976 students) did not find positive effects. He reported three experimental studies of the Accelerated Reader reading management program, which is widely used to encourage students to read books of their choice at home. Again, two of these had very small samples (50 and 39 students, respectively), and the one larger study found no differences.

COOPERATIVE LEARNING

Cooperative learning, or peer-assisted learning, refers to a set of instructional methods in which students work in small groups to help one another master academic content. Research on

the achievement effects of cooperative learning finds that cooperative methods in which there is a group goal that students can achieve only if all group members make progress increase learning across a broad range of subjects, and at grade levels from 2 to 12 (see Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003; Slavin, 1995; Slavin, Hurley, & Chamberlain, 2003; Webb & Palincsar, 1996). Specific cooperative learning methods that have been evaluated in studies that meet the inclusion criteria are as follows.

Student Teams Achievement Divisions

Student Teams Achievement Divisions (STAD; Slavin, 1995) is a simple form of cooperative learning that can be applied to any subject or grade level. Students are assigned to four-member, heterogeneous teams. They study academic materials together, and are then individually assessed. Based on the average of individual scores (or gains), teams receive recognition or other rewards. In a variation called Teams Games Tournaments, students play academic games instead of taking quizzes.

Slavin (1995) carried out a meta-analysis of experimental research on cooperative learning and found a median effect size of +0.32 across 29 studies for STAD (+0.21 for standardized tests), and +0.38 across seven studies for TGT (+0.40 for standardized tests). The majority of the studies used random assignment of teachers and/or students to cooperative learning or control groups. Many had durations shorter than the 10-week criterion used in this paper, but exceptions include four randomized studies of language arts in middle schools (Slavin, 1977, 1978, 1979; Slavin & Oickle, 1981), and of math in grade 9 classes in Philadelphia junior and senior high schools (Slavin & Karweit, 1984). Three randomized studies involving mathematics in Israel by Mevarech (1985a, 1985b, 1991) found positive effects of STAD, and even better effects if STAD was combined with mastery learning. In Nigeria, Okebukola found very positive impacts of STAD in high school science, and Hawkins et al. (1988) found positive effects of STAD for low achievers on standardized tests of reading and math but not language.

Other Cooperative Learning Methods

Several additional cooperative learning methods have also been evaluated using randomized or matched designs over periods of at least 10 weeks. A program called Reciprocal Peer Tutoring (Fantuzzo, King, & Heller, 1992) involves structured peer dyads working in specified ways on academic material. Two randomized evaluations (Fantuzzo et al., 1992; Heller & Fantuzzo, 1992) in grades 4-5 mathematics found positive effects of this approach with at-risk students. A similar approach called Classwide Peer Tutoring (Greenwood, Delquardri, & Hall, 1989; Greenwood, Terry, Utley, Montagna, & Walker, 1993) also found positive effects on reading, math, and language in a four-year implementation involving schools randomly assigned to treatments in Kansas City, Kansas. A follow-up into sixth grade found that the effects were maintained.

An extensive series of studies of cooperative learning have been carried out by David and Roger Johnson and their colleagues at the University of Minnesota (see Johnson & Johnson, 1999). Most studies of their form of cooperative learning used random assignment, but nearly all were very brief (one to six weeks). One year-long randomized study with Mexican-American children by Martinez (1990) found nonsignificant positive effects on language and math but not reading or spelling, and Kambiss (1990) found positive spelling effects.

Two additional cooperative learning methods, Reciprocal Teaching (Brown & Palincsar, 1982) and Cooperative Integrated Reading and Composition (CIRC; Stevens, Madden, Slavin, & Farnish, 1987) are discussed under “reading programs,” later in this paper.

MATHEMATICS PROGRAMS

WWC Topics Report on Middle School Mathematics

The first review published by the What Works Clearinghouse (WWC) is a “topic report” on middle school mathematics. Very few studies in this area met WWC standards, and fewer still showed positive effects of the programs evaluated.

Cognitive Tutor

Cognitive Tutor is an “intelligent tutoring system” that uses findings of research on artificial intelligence to help middle school students learn mathematics. Developed at Carnegie Mellon University, the program is disseminated by Carnegie Learning.

One randomized experiment by Morgan & Ritter (2002) met the WWC standards. It compared 360 students randomly assigned either to the Cognitive Tutor or to a textbook program. Differences significantly favored the Cognitive Tutor students (ES= +0.23).

I Can Learn

I Can Learn is a computerized algebra curriculum designed to help diverse, inner-city students succeed in algebra. An experiment by Kirby (2004) randomly assigned 254 students to I Can Learn or control treatments. The I Can Learn students gained significantly more than controls (ES=+0.41). Two matched studies that met the WWC standards obtained mixed, nonsignificant results for this program, however.

Other Middle School Programs

The What Works Clearinghouse also presented data on three additional middle school math programs. A randomized evaluation of The Expert Mathematician found effects that directionally favored the program but were not statistically significant. Two studies of Saxon Math, one randomized and one matched, found no differences. Three matched studies of the Connected Mathematics Project found mixed and nonsignificant effects.

The What Works Clearinghouse listed 15 middle school programs that had been evaluated in studies that did not meet WWC standards. These include such widely used technology programs as Compass Learning, PLATO, and Successmaker. It also listed 24 programs for which no studies were found, including all of the traditional textbook programs (e.g., Addison-Wesley, Heath, Holt, and Macmillan/McGraw-Hill, and Scott Foresman). The University of Chicago School Mathematics Program, perhaps the most widely used reform-oriented model, was listed as being still under review.

Elementary Mathematics Programs

The What Works Clearinghouse has not yet reviewed mathematics programs at the elementary level, but there are a few elementary programs with strong research bases.

Cognitively Guided Instruction (CGI)

Cognitively Guided Instruction (CGI; Carey, Fennema, Carpenter, & Franke, 1993; Carpenter, Fennema, Peterson, Chiang, & Loef, 1989) is a mathematics program designed to develop student problem solving in the early elementary grades. CGI was created to teach the teachers of first-grade students about problem-solving processes that their students use when solving simple arithmetic and complex mathematics problems and to train the teachers to create curricula consistent with new understandings of how children learn. Following extensive training, CGI teachers create units and themes to last the entire school year.

In an evaluation of CGI (Carpenter et al., 1989), 40 teachers were randomly assigned to either a control or a treatment group. Teachers in both of the groups were involved in problem-solving workshops, but one was a CGI workshop and the other was a generic problem-solving workshop. On the Iowa Test of Basic Skills (ITBS), CGI students outscored their control group counterparts in computations and in problem solving that involved complex addition/subtraction.

A second study of CGI evaluated the effectiveness of the program among low-income minority students (Villaseñor & Kepner, 1993). Twelve experimental and 12 control teachers were randomly assigned to CGI and control classes in Milwaukee. A 14-item arithmetic word-problem test focusing on higher-level cognitive processes (Carpenter et al., 1989), developed by the creators of CGI, was administered as a pretest and again as a posttest. Controlling for small pretest differences, the experimental students significantly outscored their control-group counterparts.

Project SEED

Project SEED (Hollins, Smiler, & Spencer 1994; Phillips & Ebrahimi, 1993; Project SEED, 1995) is an enrichment mathematics program designed to teach elementary school students, particularly low-income and minority students, to develop confidence in their ability to be successful in all academic work, giving them the grounding to help them to face challenging academic situations.

Project SEED hires and trains mathematicians, scientists, and engineers to teach students. They are trained to introduce abstract mathematical concepts using a discovery method based on Socratic questioning, always making students active participants in the lessons. The Project SEED curriculum does not take the place of the regular mathematics curriculum, but is a supplement to it. When the Project SEED mathematics specialists teach the students, the regular classroom teachers remain in the classroom and observe and participate in what is being taught. Students involved in the program are expected to learn using dialogue, choral responses, discussion, and debates. The Project SEED mathematics specialists also conduct workshops and do observations with the regular classroom teachers.

A study that evaluated the effects of one semester of Project SEED in Detroit (Webster & Chadbourn, 1992) compared the California Achievement Test (CAT) scores of 244 fourth grade students in SEED classrooms to those of 244 fourth grade students in SEED schools, but not in SEED classrooms (non-SEED), and to those of 244 fourth grade students neither in SEED

schools nor in SEED classrooms (comparison group). Students in all three groups were matched based on gender, ethnicity, free or reduced lunch status, and third grade CAT scores.

The SEED students outscored comparison group students in total math scores (ES=+.37), math computation (ES=+.38), and math concepts (ES=+.32). Students in the SEED group also outperformed students in the non-SEED groups on math total (ES=+.19), math computation (ES=+.16) and math concepts (ES=+.19).

READING PROGRAMS

The What Works Clearinghouse is scheduled to review research on beginning reading programs in summer 2005. As is the case in mathematics, there are relatively few programs that have even a single study evaluating their effectiveness in comparison to control groups.

Success for All

The Success for All reading program has been evaluated in more than 50 matched experimental comparisons and one randomized experiment involving 41 elementary schools across the U.S. This research was discussed earlier under “comprehensive school reform.”

Direct Instruction

The Direct Instruction reading program has been extensively evaluated and found to be effective, most notably in the federal Follow Through research (see Adams & Engelmann, 1996). Research on DI was discussed earlier under “comprehensive school reform.”

Reciprocal Teaching

Reciprocal Teaching (Palincsar & Brown, 1984) is a professional development program designed to improve the reading comprehension of children in elementary and middle schools that emphasizes cognitive strategies of scaffolding through dialogue.

The main two components of Reciprocal Teaching are comprehension fostering, which includes the four strategies of question generation, summarization, prediction, and clarification; and dialogue, which includes prepared conversations and questions that guide the comprehension process and product. The program uses a scaffolding process, in which teachers are initially more responsible for producing questions, guiding the dialogue, and showing the students how to comprehend text. Eventually, the students become more responsible for the products, creating questions for each other and guiding the dialogue with less teacher input.

A meta-analysis of the achievement effects of Reciprocal Teaching was carried out by Rosenshine & Meister (1994). Sixteen studies representing different levels of implementation (high, medium, and low) and different methods of teaching were synthesized. High implementation studies included dialogue, questions, and assessment of student learning strategies, medium level studies included dialogue but did not include assessments, and low level studies had neither dialogue nor assessment information.

The meta-analysis investigated how Reciprocal Teaching students performed on standardized and experimenter-made tests as compared to their control-group peers. The overall

effect size for performance on standardized tests was $+0.32$; but only in two cases did the Reciprocal Teaching students do significantly better on standardized tests than their control group counterparts. Effect sizes were much higher on the experimenter-made tests ($ES=+0.88$).

Cooperative Integrated Reading and Composition (CIRC)

Cooperative Integrated Reading and Composition, or CIRC (Stevens, Madden, Slavin, & Farnish, 1987), used in grades 2-8, is a cooperative learning program that involves a series of activities derived from research on reading comprehension and writing strategies. Students work in four-member heterogeneous learning teams. After the teacher introduces a story from a basal text or trade book, students work in their teams on a prescribed series of activities relating to the story. These include partner reading, where students take turns reading to each other in pairs; “treasure hunt” activities, in which students work together to identify characters, settings, problems, and problem solutions in narratives; and summarization activities. Students write “meaningful sentences” to show the meaning of new vocabulary words, and write compositions that relate to their reading. The program includes a curriculum for teaching main idea, figurative language, and other comprehension skills, and includes a home reading and book report component. The writing/language arts component of CIRC uses a cooperative writing process approach in which students work together to plan, draft, revise, edit, and publish compositions in a variety of genres. Students master language mechanics skills in their teams, and these are then added to editing checklists to ensure their application in students’ own writing. Teams earn recognition based on the performance of their members on quizzes, compositions, book reports, and other products (see Madden, Slavin, Farnish, Livingston, Calderón, & Stevens, 1996).

The original CIRC program has been evaluated in three matched studies in elementary schools (Stevens, Madden, Slavin, & Farnish, 1987; Stevens & Slavin, 1995) and one study in two middle schools (Stevens & Durkin, 1992). In each case, CIRC students made significantly greater gains than control students on standardized tests of reading achievement. Two studies in Israel, one in Hebrew and one in Arabic, also found positive effects of CIRC compared to traditional methods (Hertz-Lazarowitz et al., 1996; Schaedel et al., 1996).

Bilingual CIRC, or BCIRC (Calderón, Hertz-Lazarowitz, & Slavin, 1998; Calderón, Tinajero, & Hertz-Lazarowitz, 1992), adds to the CIRC structure several adaptations to make it appropriate to bilingual settings. It is built around Spanish reading materials in the younger grades and then uses transitional reading materials as students begin to transition from Spanish to English. In addition, effective ESL strategies designed to engage students in negotiating meaning in two languages and increase authentic oral communication are built into the training program.

A four-year study of BCIRC was conducted in 24 bilingual grade 2-4 classes in El Paso, Texas (Calderón, Hertz-Lazarowitz, & Slavin, 1998). Experimental and control classes were carefully matched. Students transitioned from mostly-Spanish instruction in second grade to mostly-English instruction in the fourth grade. BCIRC students scored significantly higher than controls on Spanish standardized reading measures in second grade and then on English standardized reading measures in third and fourth grades.

TUTORING PROGRAMS IN READING

Reading Recovery/Descubriendo La Lectura

Reading Recovery (RR; Pinnell, DeFord, & Lyons, 1988) is an early intervention tutoring program for young readers who are experiencing difficulty in their first year of reading instruction. RR serves the lowest achieving readers (lowest 20 percent) in first grade classes by providing the children with supplemental tutoring in addition to their regular reading classes. Children participating in RR receive daily one-to-one 30 minute lessons for 12-20 weeks with a teacher trained in the RR method. The lessons consist of a variety of experiences designed to help children develop effective strategies for reading and writing. When the student reaches a stage at which he or she is able to read at or above the average class level and can continue to read without later remedial help, the student is discontinued from the program. Students who are not discontinued are excluded from the program after 60 lessons and may be placed either in special education classes or in some other form of remedial education.

RR tutors are certified teachers who receive a year's training in Reading Recovery tutoring. The tutoring model emphasizes "learning to read by reading" (Pinnell, 1989; Pinnell, DeFord, & Lyons, 1988). The lessons are one-to-one tutorial sessions that include reading known stories, reading a story that was read once the day before, writing a story, working with a cut-up sentence, and reading a new book.

Descubriendo La Lectura (DLL) is a Spanish adaptation of Reading Recovery (RR), developed and studied in Tucson, Arizona. It is equivalent in all major program aspects to the original program. Students in Spanish bilingual classes whose reading scores fall in the bottom 20 percent in the first grade are placed in DLL.

The Ohio State group has conducted three randomized longitudinal studies comparing English Reading Recovery to traditional Title 1 pull-out or in-class methods. The first (pilot) study (Huck & Pinnell, 1986, Pinnell, 1988) of RR involved first grade students from six inner-city Columbus, Ohio, schools and six matched comparison classes. The lowest 20 percent of students in each class served as the experimental and control group, respectively. The second longitudinal study (Pinnell, Short, Lyons, & Young, 1986; DeFord, Pinnell, Lyons, & Young, 1988) involved 32 teachers in 12 schools in Columbus. Again, students in the lowest 20 percent of their classes were randomly assigned to Reading Recovery or control conditions. Results showed that Reading Recovery students substantially outperformed control students on almost all measures in a series of assessments developed by the program, except tests of letter identification and word recognition, both of which had ceiling effects. With the exception of these, the effects ranged from $+.57$ to $+.72$.

An oral reading measure called Text Reading Level was given at the end of first, second, and third grades. On this test, students were asked to read books that got progressively more difficult. The results of this study for Text Reading Level at the end of first grade showed substantial positive effects for both the pilot cohort and the second cohort ($ES=+.72$ and $+.78$ respectively). On a follow-up assessment at the end of second grade the effects diminished ($ES=+.29$ and $+.46$ respectively). At the end of third grade, the effect sizes had diminished even further ($ES=+.14$ and $+.25$, respectively). The raw experimental-control differences remained

about the same over the three-year period, but due to the increasing standard deviations the effect sizes diminished (see Wasik & Slavin, 1993).

A third study of Reading Recovery involved schools in ten districts throughout the state of Ohio (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). This study compared schools randomly assigned to use Reading Recovery, to one of three program variations, or to a control group. On mid-year assessments, Reading Recovery students scored better than control students and better than an RR variation that involved a shorter training period, a group (not one-to-one) version of RR, and an alternative tutoring model. A Gates-McGinitie test given in May of first grade showed small and nonsignificant effects, but the following fall RR students scored significantly higher than controls on both Text Reading Level and a dictation test. None of the RR variations were significantly higher than control groups on these measures.

Studies of Reading Recovery conducted by researchers who are not associated with the program find patterns of results similar to those found by the Ohio State researchers. Tests given immediately after the Reading Recovery intervention show substantial positive effects of the program. These effects diminish in size in years after first grade, although some difference is usually still detectable in third grade (Baenen, Bernholz, Dulaney, Banks, & Willoughby, 1995; Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Shanahan & Barr, 1995).

An evaluation of Descubriendo La Lectura (DLL) was conducted by Escamilla (1994) in Tucson. The experiment compared 23 DLL students to 23 matched comparison students in a school that did not have DLL. In both cases, students were identified as being in the lowest 20 percent of their classes based on individually administered tests and teacher judgment.

The outcomes of DLL on Spanish reading measures given at the end of first grade were extremely positive. On six scales of a Spanish Observation Survey adapted from the measures used in evaluations of the English Reading Recovery program, DLL students started out below controls and ended the year substantially ahead of them, with effect sizes (adjusted for pretest differences) ranging from +0.97 to +1.71.

Other Tutoring Programs

While Reading Recovery is by far the most extensively researched and widely used tutoring model, there are several others that have also been evaluated. Fashola, Diggs, & Smith (2003) evaluated an after school tutoring program based on the Success for All beginning reading model, and found positive effects in a randomized experiment. Early Steps, a tutoring model similar to Reading Recovery, was successfully evaluated in two yearlong matched studies with first grade children (Morris et al., 2000; Santa & Hoiem, 1999). The Howard Street Tutoring Program, which relies on volunteers, was also found to increase reading achievement in a yearlong matched experiment (Morris, Shaw, & Perney, 1990). A two-year longitudinal randomized evaluation found positive effects for a volunteer tutoring program called SMART, developed at the University of Oregon (Baker, Gersten, & Keating, 2000). Book Buddies, another volunteer model developed at the University of Virginia, was also found to increase student achievement in matched studies (Invernizzi, Rosemary, Juel, & Richards, 1997).

READING PROGRAMS FOR ENGLISH LANGUAGE LEARNERS

Slavin & Cheung (2004) recently carried out a comprehensive review of research on reading programs for English language learners and other language minority students. For beginning reading, the review found that the best-evaluated programs for English proficient students were, with appropriate adaptations, the same programs found to be effective with ELLs. Across five matched studies of an English language development adaptation of Success for All (Nunnery et al., 1997; Livingston & Flaherty, 1997; Slavin & Madden, 1999; Ross et al., 1998; Hurley et al., 2001), the median effect size on English reading measures was +0.37. Two studies of a Spanish bilingual adaptation found an effect size of +0.41 (Livingston & Flaherty, 1997; Nunnery et al., 1997). Two matched studies of Direct Instruction (Becker & Gersten, 1982; Gersten, 1985), plus one randomized study of a small group tutorial that used DI content (Gunn et al., 2000), found significant positive effects on English measures. Finally, a study of a Spanish adaptation of the Reading Recovery tutoring model found substantial positive effects (Escamilla, 1994).

For the upper elementary grades, Slavin & Cheung (2004) found a broader range of effective programs used with ELLs. These included Bilingual Cooperative Integrated Reading and Composition (BCIRC), discussed earlier, and an enriched program for Spanish-to-English transition evaluated by Saunders (1998) and Saunders & Goldenberg (1999). Carlo et al. (2004) successfully evaluated another enriched transition program. All of these emphasized cooperative learning, vocabulary development, and methods designed to encourage students to use their growing English language skills in a variety of contexts.

Dropout Prevention Programs

Two programs primarily designed to increase the high school graduation rates of at-risk students met the standards of this review: The Coca-Cola Valued Youth Program (VYP) and ALAS (Achievement for Latinos through Academic Success).

The Coca-Cola Valued Youth Program

The Coca-Cola Valued Youth Program (1991) is a cross-age tutoring program designed to increase the self-esteem and school success of at-risk middle and high school students by placing them in positions of responsibility as tutors of younger elementary school students. The Valued Youth Program was originally developed by the Intercultural Development Research Association in San Antonio, Texas, funded by Coca-Cola.

The overall goal of the program is to reduce the dropout rates of at-risk students by improving their self-concepts and academic skills. This is done by making them tutors, and providing them with assistance with basic academic skills. The program also emphasizes elimination of non-academic and disciplinary factors that contribute to dropping out. For example, it attempts to develop students' senses of self-control, decrease student truancy, and reduce disciplinary referrals. It also seeks to form home-school partnerships to increase the level of support available to students.

The tutors are required to enroll in a special tutoring class, which allows them to improve their own basic academic skills as well as their tutoring skills. The students who are involved as tutors are paid a minimum wage stipend. The tutors work with three elementary students at a

time for a total of about four hours per week. Functions are held to honor and recognize the tutors as role models. They receive t-shirts, caps, and certificates of merit for their efforts.

The main evaluation of the Coca-Cola Valued Youth Program compared 63 VYP tutors to 70 students in a comparison group (Cardenas, Montecel, Supik, & Harris, 1992). The students in four San Antonio schools were matched on the basis of age, ethnicity, lunch eligibility, percentage of students retained in grade, and scores on tests of reading, quality of school life, and self-concept. They were selected (not randomly) into the experimental group based on scheduling and availability, and then the remaining students were placed into the comparison group. Nearly all students in both groups were Latino and limited English proficient.

Two years after the program began, 12 percent of the comparison students but only 1 percent of the VYP students had dropped out. Reading grades were significantly higher for the VYP group, as were scores on a self-esteem measure and on a measure of attitude towards school.

Achievement for Latinos Through Academic Success (ALAS)

Achievement for Latinos through Academic Success (ALAS; Larson & Rumberger, 1995) is a dropout prevention program for high-risk middle school Latino students, particularly Mexican-American students from high-poverty neighborhoods. This program focuses on youth with learning and emotional/behavioral disabilities using a collaborative approach across multiple spheres of influence: home, school, and community. Students are provided with social problem solving training, counseling, and recognition for academic excellence. Family strategies include use of community resources, parent training in school participation, and training to guide and monitor adolescents. Parents are offered workshops on school participation and teen behavior management. The program also focuses on integrating school and home needs with community services, and advocating for the student and parent when necessary. Community strategies include enhancement of collaboration among community agencies for youth and family services, and enhancement of skills and methods for serving the youth and family.

ALAS was evaluated in a junior high school that was 96 percent Latino. Students who experienced ALAS had dropout rates substantially lower than a matched control group.

EVIDENCE-BASED REFORM: POLICY IMPERATIVES

The evidence cited in this paper shows both the potential of evidence-based reform and the distance yet to go. We do not know nearly as much as we'd want to know about replicable programs that can accelerate the achievement of all students. There is a particular need for more randomized, large scale studies done over meaningful periods of time to evaluate promising programs. Yet at the same time, it is clear that we are not using what we do know. None of the programs found to be effective in replicated research are used widely enough to make a difference at the policy level. The largest, Success for All, is used in about 1,400 schools, out of perhaps 70,000 elementary schools or 20,000 Title I schoolwide programs.

Educational policies needed to move toward evidence-based reform are described in the following sections.

1. Substantially Increase Support for Research and Development

Given its potential importance, funding for research in education is shockingly low, and much of what is spent goes for routine data collection, technical assistance, and other activities that are not research. It is difficult to separate out, but funding for development, evaluation, and dissemination of programs and practices for K-12 education is surely less than \$100 million per year across all agencies, and may be less than \$50 million. In contrast, the U.S. Department of Education spends about \$1 billion per year just on support for after school programs. For half this amount, \$500 million per year, researchers and developers could substantially advance knowledge about practical, effective programs for all types of schools.

2. Fund development of new programs

There are still too few promising programs in the pipeline. Researchers and developers need funding to develop new replicable programs based on current understandings of how children learn, current technologies, and current needs. In particular, the Department of Education should hold “design competitions” in which developers are challenged to create programs of all kinds to solve central problems of American education: reading, math, science, and social studies programs at all grade levels, programs for English language learners, solutions for children with reading disabilities, dropout prevention programs, school-to-work programs, classroom management programs, assessment methods, schoolwide reform models for elementary and secondary schools, and much more. In each case, a number of developers should be supported to create, pilot, and ultimately evaluate promising models. As the work progresses, additional projects should be added and ones that are not working as hoped winnowed out. As part of the development process, there is a need for basic research, including correlational, descriptive, and small-scale experimental research, to provide a base for development of research-based programs.

3. Fund evaluation of existing and new programs

IES is now funding an impressive array of rigorous evaluations of educational programs and practices, but much more remains to be done. Ideally, developers should have funding to do their own evaluations as they are preparing to scale up their programs, but then independent, third-party evaluators need support to do their own high-quality evaluations. In today’s context, a “rigorous evaluation” means one in which schools are assigned at random to use a given program or to use an alternative control program, with measurement of achievement at pre- and post-test and of implementation throughout the experiment.

4. Provide incentives for schools to participate in research

One cost-effective way to carry out randomized experiments evaluating educational programs would be to provide a competitive preference in school funding programs for schools willing to be assigned at random to use a given program immediately or one year later. For example, schools applying for funding to implement comprehensive school reform programs, secondary reading programs under the new “Striving Readers” initiative, K-3 reading programs under Reading First, or after-school programs under 21st Century Community Learning Centers, would be given a better chance of success if they agree to participate in randomized evaluations. In this way, the cost of the research would just be the data collection and analysis, not the

program implementation (which usually consumes the majority of the funding for randomized field research).

5. Provide incentives for schools and districts to use programs validated in rigorous research

As this paper illustrates, there are already many programs with strong evidence of effectiveness, and many more will be validated in the coming years. District and school leaders need clear information on the findings of this research, but information alone will not suffice, as large textbook and technology companies will continue to demonstrate that marketing is more powerful than evidence. Yet the federal government has many levers it can use to promote adoption of programs validated in rigorous research. For example, it can give competitive preferences to schools that use proven programs in discretionary grants. It can insist that schools failing to meet adequate yearly progress (AYP) standards for three years choose a proven model. It can ask schools not meeting AYP standards to explain in their Title I plans why they feel it is important to continue to use programs with no evidence of effectiveness when well-validated alternatives exist. The necessary language for this policy already resides in No Child Left Behind, but it would be necessary to redefine programs and practices “based on scientifically-based research” as ones that have been evaluated in comparison to control groups and found to be effective in increasing student achievement.

6. Maintain the integrity of proven programs

If evidence-based programs are emphasized in educational policy, there needs to be some oversight to ensure that the programs being adopted are essentially the same as the ones that were proven to be effective. For example, a program that was successful with extensive training and follow-up could not be considered “evidence-based” if it were later disseminated with minimal professional development.

7. Encourage states to base policies on research

So far, the evidence-based policy movement has been almost entirely a federal initiative. State departments of education need to embrace this dynamic if it is to take hold on a wide scale. For example, the states now control a fund that amounts to 4 percent of their Title I funding to help schools meet adequate yearly progress standards. No Child Left Behind encourages them to use this money to help schools adopt proven programs, but this is unlikely to happen unless the states dedicate themselves to evidence-based policies and the Department of Education monitors how states use the 4 percent set aside.

CONCLUSION

The solutions to America’s educational problems must draw on our nation’s greatest strength, the ingenuity, inventiveness, and technological capacity of the American people. America leads the world in medicine, agriculture, and technology because of its unequalled national capacity to create new solutions. This dynamic has not taken hold in education, but there is no reason that it cannot do so. With a modest investment of, say, \$500 million per year in R&D, our country can bring about a revolution in education. This will be beneficial to all children, but especially to those who are least well served by today’s schools.

References

- Adams, G.L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.
- Amrein, A., & Berliner, D. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, 60 (5), 32-38.
- Baenen, N., Bernholc, A., Dulaney, C., Banks, K., Willoughby, M. (1995). *Evaluation report: WCPSS Reading Recovery 1990-1994*. Raleigh, NC: Wake County Public Schools.
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, 35 (4), 494-519.
- Bangert-Drowns, R.L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69-93.
- Bebell, D., O'Dwyer, L., Russell, M., & Seeley, K. (2004). *Estimating the effect of computer use at home and in school on student achievement*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Becker, H. J. (2001, April). *How are teachers using computers in instruction?* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Becker, W.C., & Gersten, R. (1982). A follow-up on Follow Through: The later effects of the Direct Instruction model on children in fifth and sixth grades. *American Educational Research Journal*, 19 (1), 75-92.
- Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research*, 72(1), 101-130.
- Borman, G., & Hewes, G. (2003). Long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, 24 (2), 243-266.
- 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 (2), 125-230.
- Borman, G.D., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (in press). Success for All: First year results from the National Randomized Field Trial. *Educational Evaluation and Policy Analysis*.
- Brown, A., & Palincsar, A. (1982). Inducing strategic learning from text by means of informed, self-controlled training. *Topics in Learning and Learning Disabilities*, 2, 1-17.

- Calderón, M., Hertz-Lazarowitz, R., & Slavin, R.E. (1998). Effects of Bilingual Cooperative Integrated Reading and Composition on students making the transition from Spanish to English reading. *Elementary School Journal*, 99 (2), 153-165.
- Calderón, M., Tinajero, J., & Hertz-Lazarowitz, R. (1992). Adapting CIRC to meet the needs of bilingual students. *Journal of Educational Issues of Linguistic Minority Students*, 10, 79-106.
- Campbell, M., Cousins, E., Farrell, G., Kamii, M., Lam, D., Rugen, L. & Udall, D. (1996). The Expeditionary Learning Outward Bound design. In S. Stringfield, S. Ross, & L. Smith (Eds.), *Bold plans for school restructuring: The New American Schools designs*. Mahwah, NJ: Erlbaum.
- Cardenas, J. A., Montecel, M.R., Supik, J.D. & Harris, R.J. (1992). The Coca-Cola Valued Youth Program: Dropout prevention strategies for at-risk students. *Texas Researcher*, 3, 111-130.
- Carey, D.A., Fennema, E., Carpenter, T.P., & Franke, M.L. (1993). Equity and mathematics education. In W. Secada, E. Fennema, & L. Byrd (Eds.), *New directions in equity for mathematics education*. New York: Teachers College Press.
- Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D., Lively, T., & White, C. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39 (2), 188-215.
- Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C. P., Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26 (4), 499-531.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24 (4), 305-331.
- Center on Education Policy (2003). *State and federal efforts to implement the No Child Left Behind Act*. Washington, DC: Author.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30, 240-261.
- Clay, M.M. (1985). *The early detection of reading difficulties*. Exeter, NH: Heinemann.
- Coca-Cola Valued Youth Program (1991). *Proposal submitted to the Program Effectiveness Panel of the U.S. Department of Education*. Washington, DC: U.S. Department of Education.
- Cochran-Smith, M. (1991). Word processing and writing in elementary classrooms: A critical review of related literature. *Review of Educational Research*, 61(1), 107-155.

- 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 (3), 543-597.
- Cook, T.D., Murphy, R.F., & Hunt, H.D. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, 37 (2), 535-597.
- Darling-Hammond, L. (1995). Inequality and access to knowledge. In J. Banks & C. A. M. Banks (Eds.), *Handbook of Research on Multicultural Education* (pp. 465-483). New York: Macmillan.
- DeFord, Pinnell, G.S., Lyons, C., & Young, P. (1988). *Reading Recovery: Volume IX, report on the follow-up studies*. Columbus, OH: Ohio State University.
- Escamilla, K. (1994). Descubriendo La Lectura: An early intervention literacy program in Spanish. *Literacy, Teaching, and Learning*, 1(1), 57-70.
- Fantuzzo, J.W., King, J.A., & Heller, L.R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, 84 (3), 331-339.
- Fashola, O., Diggs, W., and Smith, D. (2003). *Implementation and Evaluation of a One-to-One Tutorial Program in Baltimore City Public Schools*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk, Johns Hopkins University.
- Gawande, A. (2004). The bell curve. *The New Yorker*, Dec. 6, 2004, 82-91.
- Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis*, 7 (3), 187-196.
- Greenwood, C.R., Delquadri, J.C., & Hall, R.V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, 81, 371-383.
- Greenwood, C.R., Terry, B., Utley, C. A., Montagna, D., & Walker, D. (1993). Achievement, placement, and services: Middle school benefits of Classwide Peer Tutoring used at the elementary level. *School Psychology Review*, 22 (3), 497-516.
- Grigg, W., Daane, M., Jin, Y., Campbell, J. (2003). *The nation's report card: Reading 2002*. Washington, DC: US Department of Education.
- Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education*, 34 (2), 90-103.

- Hawkins, J.D., Doueck, H.J., & Lishner, D.M. (1988). Changing teacher practices in mainstream classrooms to improve bonding and behavior of low achievers. *American Educational Research Journal*, 25 (10), 31-50.
- Heady, R., & Kilgore, S. (1996). The Modern Red Schoolhouse. In S. Stringfield, S. Ross, & L. Smith (Eds.), *Bold plans for school restructuring: The New American Schools designs*. Mahwah, NJ: Erlbaum.
- Heller, L.R., & Fantuzzo, J.W. (1992). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review*, 64-105.
- Herman, R. (1999). *An educator's guide to schoolwide reform*. Arlington, VA: Educational Research Service.
- Hertz-Lazarowitz, R., Lerner, M., Schaedel, B., Walk, A., & Sarid, M. (1996). Story-related writing: An evaluation of CIRC in Israel. *Helkat-Lashon* (Journal of Linguistic Education, in Hebrew).
- Hollins, E.R., Smiler, H., & Spencer, K. (1994). Benchmarks in meeting the challenges of effective schooling for African American youngsters. In E.R. Hollins, J.E. King, & W.C. Hayman (Eds.), *Teaching diverse populations: Formulating a knowledge base*. Albany: State University of New York Press.
- Huck, C.S., & Pinnell, G.S. (1986). *The Reading Recovery project in Columbus, Ohio: Pilot year, 1984-85*. Columbus, OH: Ohio State University.
- Hurley, E.A., Chamberlain, A., Slavin, R.E., & Madden, N.A. (2001). Effects of Success for All on TAAS Reading: A Texas statewide evaluation. *Phi Delta Kappan*, 82 (10), 750-756.
- Invernizzi, M., Rosemary, C., Juel, C., & Richards, H. (1997). At-risk readers and community volunteers: A three-year perspective. *Journal of Scientific Studies of Reading*, 1, 277-300.
- Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Boston: Allyn & Bacon.
- Kaestle, C.F. (1993). The awful reputation of educational research. *Educational Researcher*, 22 (1), 23-26-31.
- Kambiss, P.A. (1990). *The effects of cooperative learning on student achievement in a fourth grade classroom*. Research project report, Mercer University.
- Kirby, P. (2004). *Comparison of I Can Learn and traditionally-taught 8th grade student performance on the Georgia Criterion-Referenced Competency Test*. Unpublished manuscript.

- Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. SRI Project Number P10446.001. Arlington, VA: SRI International.
- Larson, K., & Rumberger, R. (1995). Doubling school success in highest-risk Latino youth: Results from a middle school intervention study. In R. F. Macias & R. Garcia Ramos (Eds.), *Changing schools for changing students*. Santa Barbara, CA: University of California at Santa Barbara.
- Levin, H.M. (1987). Accelerated schools for disadvantaged students. *Educational Leadership*, 44(6), 19-21.
- Livingston, M. & Flaherty, J. (1997). *Effects of Success for All on reading achievement in California schools*. Los Alamitos, CA: WestEd.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449-521.
- Madden, N.A., Slavin, R.E., Farnish, A.M., Livingston, M.A., & Calderón, M. (1999). *Reading Wings teachers' manual*. Baltimore, MD: Success for All Foundation.
- Martinez, L.J. (1990). *The effect of cooperative learning on academic achievement and self-concept with bilingual third-grade students*. Unpublished doctoral dissertation, United States International University.
- May, H., Supovitz, J.A., & Perda, D. (2004). *A longitudinal study of the impact of America's Choice on student performance in Rochester, New York, 1998-2003*. Philadelphia, PA: Consortium for Policy Research in Education.
- Mevarech, Z.R. (1985a). The effects of cooperative mastery learning strategies on mathematics achievement. *Journal of Educational Research*, 78, 372-377.
- Mevarech, Z.R. (1985b, April). Cooperative mastery learning strategies. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Mevarech, Z.R. (1991). Learning mathematics in different mastery environments. *Journal of Educational Research*, 84 (4), 225-231.
- Morgan, P., & Ritter, S. (2002). *An experimental study of the effects of Cognitive Tutor Algebra I on student knowledge and attitude*. (Available from Carnegie Learning, Inc., 1200 Penn Ave., #150, Pittsburgh, PA 15222)
- Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in grades two and three: An after-school volunteer tutoring program. *Elementary School Journal*, 91, 133-150.

- Morris, D., Tyner, B., & Perney, J. (2000). Early steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, 92, 681-693.
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in educational research*. Washington, DC: Brookings.
- National Center for Education and the Economy (2003). *America's Choice: Program overview*. Washington, DC: Author. Available at www.ncee.org/acsd.
- National Commission on Excellence in Education (1983). *A nation at risk*. Washington, DC: U.S. Department of Education.
- Neill, M., & Gaylor, K. (2001). Do high-stakes graduation tests improve learning outcomes: Using state-level NAEP data to evaluate the effects of mandatory graduation tests. In G. Orfield & M.L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 107-126). New York: Century Foundation Press.
- Nunnery, J., Ross, S., Smith, L., Slavin, R., Hunter, P., & Stubbs, J. (1997, March). *Effects of full and partial implementations of Success for All on student reading achievement in English and Spanish*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- OECD (2004). *First results from PISA 2003*. Paris: Author. Available at www.pisa.oecd/dataoecd.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction*, 2, 117-175.
- Phillips, H., & Ebrahimi, H. (1993). Equation for success: Project SEED. In G. Cuevas & M. Driscoll (Eds.), *Reaching all students with mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Pinnell (1988, April). *Sustained effects of a strategy-centered early intervention program in reading*. Paper Presented at the Annual Convention of the American Educational Research Association, New Orleans.
- Pinnell, G.S. (1989). Reading Recovery: Helping at-risk children learn to read. *Elementary School Journal*, 90, 161-182.
- Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). *Reading Recovery: Early intervention for at-risk first graders*. Arlington, VA: Education Research Service.
- Pinnell, G.S., Lyons, C.A., DeFord, D.E., Bryk, A.S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 9-40.

- Pinnell, G.S., Short, A.G., Lyons, C.A., & Young, P. (1986). *The Reading Recovery project in Columbus, OH, Year I: 1985-1986*. Columbus, OH: Ohio State University.
- Project SEED, Inc. (1995). *Project SEED: Submission to the program effectiveness panel of the U.S. Department of Education*. Berkeley, CA & Dallas, TX: Author.
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 94* (2), 240-257.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 479-530.
- Ross, S.M., Sanders, W.L., & Wright, S.P. (1998). *An analysis of Tennessee Value Added Assessment (TVAAS) performance outcomes of Roots and Wings schools from 1995-1997*. Memphis: University of Memphis.
- Santa, C., & Hoiem, T. (1999). An assessment of Early Steps: A program for early intervention of reading problems. *Reading Research Quarterly, 34*, 54-79.
- Saunders, W.M. (1998). *Improving literacy achievement for English learners in transitional bilingual programs*. Long Beach, CA: Center for Research on Education, Diversity, and Excellence, University of California.
- Saunders, W.M., & Goldenberg, C. (1999). *The effects of a comprehensive language arts transition program on the literacy development of English language learners*. Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence, University of California.
- Schaedel, B., Hertz-Lazarowitz, R., Walk, A., Lerner, M., Juberan, S., & Sarid, M. (1996). The Israeli CIRC (ALASH): First year achievements in reading and comprehension. *Helkat-Lashon* (Journal of Linguistic Education, in Hebrew), *23*, 401-423.
- Shanahan, T., & Barr, R. (1995). *Reading Recovery: An independent evaluation of an early instructional intervention for at risk learners*. Chicago: University of Illinois at Chicago.
- Slavin, R.E. (1977). A student team approach to teaching adolescents with special emotional and behavioral needs. *Psychology in the Schools, 14* (10), 77-84.
- Slavin, R.E. (1978). Student teams and achievement divisions. *Journal of Research and Development in Education, 12*, 39-49.
- Slavin, R.E. (1979). Effects of biracial learning teams on cross-racial friendships. *Journal of Educational Psychology, 71*, 381-387.

- Slavin, R.E. (1995). *Cooperative learning: Theory, research, and practice* (2nd Ed.). Boston: Allyn & Bacon.
- Slavin, R.E. (2003). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31 (7), 15-21.
- Slavin, R.E., & Cheung, A. (2004). *Effective reading programs for English language learners*. Manuscript submitted for publication.
- Slavin, R.E., Hurley, E.A., & Chamberlain, A.M. (2003). Cooperative learning and achievement: Theory and research. In W.M. Reynolds & G.E. Miller (Eds.), *Handbook of Psychology*, Volume 7 (pp. 177-198). Hoboken, NJ: Wiley.
- Slavin, R.E., & Karweit, N. (1984). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal*, 21, 725-736.
- Slavin, R.E., & Madden, N.A. Effects of bilingual and English as a second language adaptations of Success for All on the reading achievement of students acquiring English. *Journal of Education for Students Placed at Risk*, 1999, 4 (4), 393-416.
- Slavin, R.E., & Madden, N.A. (Eds.) (2001). *One million children: Success for All*. Thousand Oaks, CA: Corwin.
- Slavin, R.E., & Madden, N.A. (2002). Success for All and African-American and Latino students. In J. Chubb and T. Loveless (Eds.), *Bridging the achievement gap*. Washington, DC: Brookings.
- Slavin, R.E., & Madden, N.A. (Eds.) (2001). *One million children: Success for All*. Thousand Oaks, CA: Corwin.
- Slavin, R.E., Madden, N.A., Dolan, L.J., & Wasik, B.A. (1996). *Every child, every school: Success for All*. Newbury Park, CA: Corwin.
- Slavin, R.E. & Oickle, E. (1981). Effects of cooperative learning teams on student achievement and race relations: Treatment by race interactions. *Sociology of Education*, 54, 174-180.
- Stevens, R.J., Madden, N.A., Slavin, R.E., & Farnish, A.M. (1987). Cooperative Integrated Reading and Composition: Two field experiments. *Reading Research Quarterly*, 22, 433-454.
- Stevens, R.J. & Durkin, S. (1992). *Using student team reading and student team writing in middle schools: Two evaluations*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students. Report No. 36.

- Stevens, R.J., & Slavin, R.E. (1995). Effects of a cooperative approach in reading and writing on academically handicapped and nonhandicapped students. *The Elementary School Journal*, 95(3), 241-262.
- Supovitz, J., Pogliinco, S., & Snyder, B. (2001). *Moving mountains: Successes and challenges of the America's Choice comprehensive school reform design*. Philadelphia, PA: Consortium for Policy Research in Education.
- Supovitz, J., Taylor, B., & May, H. (2002). *Impact of America's Choice on student performance in Duval County, Florida*. Philadelphia, PA: Consortium for Policy Research in Education.
- Traub, J. (1999). *Better by design? A consumer's guide to schoolwide reform*. Washington, DC: Thomas Fordham Foundation.
- U. S. Department of Education (2002). *No Child Left Behind: A desktop reference*. Washington, DC: Author. Available at www.ed.gov/offices/OESE/reference.
- Villaseñor, J.R.A. Kepner, H.S. (1993). Arithmetic from a problem solving perspective: An urban implementation. *American Educational Research Journal*, 21(1) 62-69.
- Wasik, B.A., & Slavin, R.E. (1993). Preventing early reading failure with one-to-one tutoring: A best-evidence synthesis. *Reading Research Quarterly*, 28, 178-200.
- Webb, N.M., & Palincsar, A.S. (1996). Group processes in the classroom. N D.C. Berliner & R.C. Calfee (Eds.), *Handbook of Educational Psychology*. New York: Simon & Schuster Macmillan.
- Webster, W.J., & Chadbourn, R.A. (1992). *The evaluation of Project SEED*. Dallas, TX: Dallas Independent School District.
- Whitehurst, G. (2002). *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.