Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis

Robert E. Slavin
Cynthia Lake
Johns Hopkins University

-VERSION 1.2-

February, 2007

This paper was written under funding from the Institute of Education Sciences, U.S. Department of Education (Grant No. R305A040082). However, any opinions expressed are those of the authors and do not necessarily represent Department of Education positions or policies.

We thank Alan Cheung for assistance with computation of effect sizes, and Harris Cooper, Robert Boruch, Henry Becker, Philip Abrami, and Larry Suter for comments on earlier drafts.

Abstract

This article reviews research on the achievement outcomes of three types of approaches to improving elementary mathematics: Mathematics curricula, computer-assisted instruction (CAI), and instructional process programs. Study inclusion requirements included use of a randomized or matched control group, a study duration of 12 weeks, and achievement measures not inherent to the experimental treatment. Eighty-seven studies met these criteria, of which 36 used random assignment to treatments. There was limited evidence supporting differential effects of various mathematics textbooks. Effects of CAI were moderate. The strongest positive effects were found for instructional process approaches such as forms of cooperative learning, classroom management and motivation programs, and supplemental tutoring programs. The review concludes that programs designed to change daily teaching practices appear to have more promise than those that deal primarily with curriculum or technology alone.

The mathematics achievement of American children is improving, but still has a long way to go. On the National Assessment of Education Progress (2005), the math scores of fourth graders have steadily improved since 1990, increasing from 12% proficient or above to 35%. Among eighth graders, the percentage of students scoring proficient or better gained from 15% in 1990 to 20% in 2005. These trends are much in contrast to the trend in reading, which changed only slightly between 1992 and 2005.

However, while mathematics performance has grown substantially for all subgroups, the achievement gap between African American and Hispanic students and their White counterparts remains wide. In 2005, 47% of White fourth graders scored at or above "proficient" on NAEP, but only 13% of African Americans and 19% of Hispanics scored this well.

Further, the U.S. remains behind other developed nations in international comparisons of mathematics achievement. For example, U.S. 15-year-olds ranked 28th on the OECD PISA study of mathematics achievement, significantly behind countries such as Finland, Canada, the Czech Republic, and France.

While we can celebrate the growth of America's children in mathematics, we cannot be complacent. The achievement gap between children of different ethnicities, and between U.S. children and those in other countries, gives us no justification for relaxing our focus on improving mathematics for all children. Under No Child Left Behind, schools can meet their adequate yearly progress goals only if all subgroups meet state standards (or show adequate growth) in all subjects tested. Nationally, thousands of schools are under increasing sanctions because the school, or one or more subgroups, is not making sufficient progress in math. For this reason, educators are particularly interested in implementing programs and practices that have been shown to improve the achievement of all children.

One way to reduce mathematics achievement gaps, and to improve achievement overall, is to provide low-performing schools training and materials known to be markedly more effective than typical programs. No Child Left Behind, for example, emphasizes the use of research-proven programs to help schools meet their adequate yearly progress goals. Yet for such a strategy to be effective, it is essential to know what specific programs are most likely to improve mathematics achievement. The purpose of this review is to summarize and interpret the evidence on elementary mathematics programs in hopes of informing policies and practices designed to reduce achievement gaps and improve the mathematics achievement of all children.

Reviews of Effective Programs in Elementary Mathematics

Throughout the No Child Left Behind Act, there is a strong emphasis on encouraging schools to use federal funding (such as Title I) on programs with strong evidence of effectiveness from "scientifically-based research." NCLB defines scientifically-based research as research that uses experimental methods to compare programs to control groups, preferably with random assignment to conditions. Yet in mathematics, what programs meet this standard? There has never been a published review of scientific research on all types of effective programs. There have been meta-analyses on outcomes of particular approaches to mathematics education, such as use of educational technology (e.g, Becker, 1991; Chambers, 2003; Kulik, 2003; Murphy, Penuel, Means, Rorbak, Whaley, & Allen, 2002), calculators (e.g., Ellington, 2003), and math approaches for at-risk children (e.g., Baker, Gersten, & Lee, 2002; Kroesbergen & Van Luit, 2003). There have been reviews of the degree to which various math programs correspond to current conceptions of curriculum, such as those carried out by Project 2061 evaluating middle school math textbooks (AAAS, 2000). The What Works Clearinghouse (2006) is doing a review of research on effects of alternative math textbooks, but this has not yet appeared as of this writing. However, there are no comprehensive reviews of research on all of the programs and practices available to educators.

In 2002, the National Research Council convened a blue-ribbon panel to review evaluation data on the effectiveness of mathematics curriculum materials, focusing in particular on innovative programs supported by the National Science Foundation but also looking at evaluations of non-NSF materials (National Research Council, 2004). The NRC panel assembled research evaluating elementary and secondary math programs, and ultimately agreed on 63 quasi-experimental studies covering all grade levels, K-12, that they considered to meet minimum standards of quality.

The authors of the NRC (2004) report carefully considered the evidence across the 63 studies and decided that they did not warrant any firm conclusions. Using a vote-count procedure, they reported that among 46 studies of innovative programs that had been supported by NSF, 59% found significantly positive effects, 6% significant negative effects, and 35% found no differences. Most of these studies involved elementary and secondary programs of the *University of Chicago School Mathematics Project.* For commercial programs, the corresponding percentages were 29%, 13%, and 59%. Based on this, the report tentatively suggested that NSF-funded programs had better outcomes. Other than this very general finding, the report is silent about the evidence on particular programs. In addition to concerns about methodological limitations, the report maintains that it is not enough to show differences in student outcomes; curricula, they argue, should be reviewed for content by math educators and mathematicians to be sure they correspond to current conceptions of what math content should be. None of the studies combined this kind of curriculum review with rigorous evaluation methods, so the NRC chose not to describe the outcomes it found in the 63 evaluations that met its minimum standards.

Focus of the Current Review

The present review examines research on all types of math programs that are available to elementary educators today. The intention is to place all types of programs on a common scale. In this way, we hope to provide educators with meaningful, unbiased information that they can use to select programs and practices most likely to make a difference with their students. In addition, the review is intended to look broadly for factors that might underlie effective practices across programs and program types, and to inform an overarching theory of effective instruction in elementary mathematics.

The review examines three general categories of math approaches. One is mathematics curricula, in which the main focus of the reform is on introduction of alternative textbooks. Some of the programs in this category were developed with extensive funding by NSF, which began in

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

the 1990s. Such programs provide professional development to teachers, and many include innovative instructional methods. However, the primary theory of action behind this set of reforms is that higher level objectives including a focus on developing critical mathematics concepts and problem solving skills, pedagogical aids such as use of manipulatives, improved sequencing of objectives, and other features of textbooks will improve student outcomes. Outcomes of such programs have not been comprehensively reviewed previously, although they are currently being reviewed by the What Works Clearinghouse (2006).

A second major category of programs is computer-assisted instruction (CAI) programs that use technology to enhance student achievement in mathematics. CAI programs are almost always supplementary, meaning that students experience a full textbook-based program in mathematics and then go to a computer lab or a classroom-based computer to receive additional instruction. CAI programs diagnose students' levels of performance and then provide exercises tailored to students' individual needs. Their theory of action depends substantially on this individualization, and on the computer's ability to continuously assess students' progress and accommodate to their needs. This is the one category of program that has been extensively reviewed in the past, most recently by Kulik (2003), Murphy et al. (2002), and Chambers (2003).

A third category is instructional process programs. This set of interventions is highly diverse, but what characterizes its approaches is a focus on teachers' instructional practices and classroom management strategies, rather than on curriculum or technology. With two exceptions, studies of instructional process strategies hold curriculum constant. Instructional process programs introduce variations in within-class grouping arrangements (as in cooperative learning or tutoring), and in amounts and uses of instructional time. Their theories of action emphasize enhancing teachers' abilities to motivate children, to engage their thinking processes, to improve classroom management, and to accommodate instruction to students' needs. Their hallmark is extensive professional development, usually incorporating followup and continuing interactions among the teachers themselves.

The three approaches to mathematics reform can be summarized as follows:

- Change the curriculum
- Supplement the curriculum with computer-assisted instruction, or
- Change classroom practices

The categorization of programs in this review relates to a longstanding debate within research on technology by Kozma (1994) and Clark (2001). Clark (2001) argued that research on technology must hold curriculum constant, to identify the unique contributions of the technology. Kozma (1994) replied that technology and curriculum were so intertwined that it was not meaningful to separate them in analysis. As a practical matter, content, media, and instructional processes are treated in different ways in the research discussed here. The mathematics curricula vary textbooks, but otherwise do not make important changes in media or instructional methods. The CAI studies invariably consider technology and curricula together; none do as Clark (2001)

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

suggested. Most of the instructional process studies vary only the teaching methods and professional development, holding curriculum constant, but a few (*TAI Math, Project CHILD, Direct Instruction,* and *Project SEED*) combine curricula, processes, and (in the case of *Project CHILD*) media.

The categorization of the programs was intended to facilitate understanding and contribute to theory, not to restrict the review. No studies were excluded due to lack of fit within one of the categories.

This review examines research on dozens of individual programs to shed light on the broad types of mathematics reforms most likely to enhance the mathematics achievement of elementary school children.

Review Methods

This article reviews studies of elementary mathematics programs in an attempt to apply consistent, well-justified standards of evidence to draw conclusions about effective elementary mathematics programs. The review applies a technique called “best-evidence synthesis” (Slavin, 1986, 2007), which seeks to apply consistent, well-justified standards to identify unbiased, meaningful quantitative information from experimental studies, and then discusses each qualifying study, computing effect sizes but also describing the context, design, and findings of each study. Best-evidence synthesis closely resembles meta-analysis (Cooper, 1998; Lipsey & Wilson, 2001), but it requires more extensive discussion of key studies instead of primarily pooling results across many studies. In reviewing educational programs, this distinction is particularly important, as there are typically few studies of any particular program, so understanding the nature and quality of the contribution made by each study is essential. The review procedures, described below, are similar to those applied by the What Works Clearinghouse (2005, 2006). For a detailed description and justification for the review procedures used here and in syntheses by Cheung & Slavin (2005) and Slavin & Lake (2007), see Slavin, 2007.

The purpose of this review is to examine the quantitative evidence on elementary mathematics programs to discover how much of a scientific basis there is for competing claims about effects of various programs. Our intention is to inform practitioners, policymakers, and researchers about the current state of the evidence on this topic as well as identify gaps in the knowledge base in need of further scientific investigation.

Limitations of the Review

This article is a quantitative synthesis of achievement outcomes of alternative mathematics approaches. It does not report on qualitative or descriptive evidence, attitudes, or other non-achievement outcomes. These are excluded not because they are unimportant, but because space limitations do not allow for a full treatment of all of the information available on each program. Each report cited, and many that were not included (listed in Appendix 1), contain much valuable information, such as descriptions of settings, non-quantitative and non-achievement outcomes, and the story of what happened in each study. The present article extracts from these rich sources just the information on experimental-control differences on quantitative achievement measures, to contribute to an understanding of likely achievement impacts of using each of the programs discussed. Studies are included or excluded, and referred to as being "high" or "low" in quality, solely based on their contribution to an unbiased, well-justified quantitative estimate of the strength of the evidence supporting each program. For a deeper understanding of all of the findings of each study, please see the original reports.

Literature Search Procedures

A broad literature search was carried out in an attempt to locate every study that could possibly meet the inclusion requirements. This included obtaining all of the elementary studies cited by the National Research Council (2004) and by other reviews of mathematics programs, including technology programs that teach math (e.g., Chambers, 2003; Kulik, 2003; Murphy et al., 2002). Electronic searches were made of educational databases (JSTOR, ERIC, EBSCO, PsychInfo, Dissertation Abstracts), web-based repositories (Google, Yahoo, Google Scholar), and math education publishers' websites. Citations of studies appearing in the studies found in the first wave were also followed up.

Effect Sizes

In general, effect sizes were computed as the difference between experimental and control individual student posttests after adjustment for pretests and other covariates, divided by the unadjusted control group standard deviation. If the control group SD was not available, a pooled SD was used. Procedures described by Lipsey & Wilson (2001) and Sedlmeier & Gigerenzor (1989) were used to estimate effect sizes when unadjusted standard deviations were not available, as when the only standard deviation presented was already adjusted for covariates, or when only gain score SD's were available. School- or classroom-level SD's were adjusted to approximate individual-level SD's, as aggregated SD's tend to be much smaller than individual SD's. If pretest and posttest means and SD's were presented but adjusted means were not, effect sizes for pretests were subtracted from effect sizes for posttests.

Criteria for Inclusion

Criteria for inclusion of studies in this review were as follows.

1. The studies involved elementary (K-5) children, plus sixth graders if they were in elementary schools.

2. The studies compared children taught in classes using a given mathematics program to those in control classes using an alternative program or standard methods.

3. Studies could have taken place in any country, but the report had to be available in English.

4. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., analyses of covariance) had to be used. Studies without control groups, such as pre-

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

post comparisons and comparisons to "expected" gains, were excluded. Studies with pretest differences of more than 50% of a standard deviation were excluded, because even with analyses of covariance, large pretest differences cannot be adequately controlled for, as underlying distributions may be fundamentally different. See "Methodological Issues," later in this article, for a discussion of randomized and matched designs.

5. The dependent measures included quantitative measures of mathematics performance, such as standardized mathematics measures. Experimenter-made measures were accepted if they are described as comprehensive measures of mathematics, which would be fair to the control groups, but measures of math objectives inherent to the program (but unlikely to be emphasized in control groups) were excluded. For example, a study of CAI by Van Dusen & Worthen (1994) found no differences on a standardized test (ES=+0.01) but a substantial difference on a test made by the software developer (ES=+0.35). The software-specific measure was excluded, as it probably focused on objectives and formats practiced in the CAI group but not in the control group. See "Methodological Issues," later in this article, for a discussion of this issue.

6. A minimum treatment duration of 12 weeks was required. This requirement is intended to focus the review on practical programs intended for use for the whole year, rather than brief investigations. Brief studies may not allow programs intended to be used over the whole year to show their full effect. On the other hand, brief studies often advantage experimental groups that focus on a particular set of objectives during a limited time period while control groups spread that topic over a longer period. For example, a study by Cramer, Pose, & del Mas (2002) evaluated a fractions curriculum that is part of the *Rational Number Project* in a 30-day experiment. Control teachers using standard basals were asked to delay their fractions instruction to January to match the exclusive focus of the experimental group on fractions, but it seems unlikely that their focus would have been equally focused on fractions, the only skill assessed.

Appendix 1 lists studies that were considered but excluded according to these criteria, as well as the reasons for exclusion. Appendix 2 lists abbreviations used throughout the review.

Methodological Issues in Studies of Elementary Mathematics Programs

The three types of mathematics programs reviewed here, mathematics curricula, CAI programs, and instructional process programs, suffer from different characteristic methodological problems (see Slavin, 2007). Across most of the evaluations, lack of random assignment is a serious problem. Matched designs are used in most studies that met the inclusion criteria, and matching leaves studies open to selection bias. That is, schools or teachers usually choose to implement a given experimental program and are compared to schools or teachers who did not choose the program. The fact of this self-selection means that no matter how well experimental and control groups are matched on other factors, the experimental group is likely to be more receptive to innovation, more concerned about math, have greater resources for reform, or otherwise to have advantages that cannot be controlled for statistically. Alternatively, it is possible that schools that would choose a given program might be dissatisfied with their results in the past, and might

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

therefore be less effective than comparable schools. Either way, matching reduces internal validity by allowing for the possibility that outcomes are influenced by whatever (unmeasured) factors that led the school or teacher to choose the program. It affects external validity in limiting generalization of findings to schools or teachers who similarly chose to use the program.

Garden-variety selection bias is bad enough in experimental design, but many of the studies suffer from design features that add to concerns about selection bias. In particular, many of the curriculum evaluations use a post-hoc design, in which a group of schools using a given program, perhaps for many years, is compared after the fact to schools that matched the experimental program at pretest or that matched on other variables, such as poverty or reading measures. The problem is that only the “survivors” are included in the study. Schools that bought the materials, received the training, but abandoned the program before the study took place are not in the final sample, which is therefore limited to more capable schools. As one example of this, Waite (2000), in an evaluation of *Everyday Mathematics*, described how 17 schools in a Texas city originally received materials and training. Only 7 were still implementing it at the end of the year, and 6 of these agreed to be in the evaluation. We are not told why the other schools dropped out, but it is possible that the staffs of the remaining 6 schools may have been more capable or motivated than those that dropped the program. The comparison group within the same city was likely composed of the full range of more and less capable school staffs, and they presumably had the same opportunity to implement *Everyday Mathematics* but chose not to do so. Other post-hoc studies, especially those with multi-year implementations, must have also had some number of dropouts, but typically do not report how many schools there were at first and how many dropped out. There are many reasons schools may have dropped out, but it seems likely that any school staff able to implement any innovative program for several years is a more capable, more reform-oriented, or better-led staff than those unable to do so, or (even worse) than those that abandoned the program because it was not working. As an analog, imagine an evaluation of a diet regimen that only studied people who kept up the diet for a year. There are many reasons a person might abandon a diet, but chief among them is that it is not working, so looking only at the non-dropouts would bias such a study.

Worst of all, post-hoc studies usually report outcome data selected from many potential experimental and comparison groups, and may therefore report on especially successful schools using the program or matched schools that happen to have made particularly small gains, making an experimental group look better by comparison. The fact that researchers in post-hoc studies often have pre- and posttest data readily available on hundreds of potential matches, and may deliberately or inadvertently select the schools that show the program to best effect, means that readers must take results from after-the-fact comparisons with a grain of salt.

Finally, because post-hoc studies can be very easy and inexpensive to do, and are usually contracted for by publishers rather than supported by research grants or done as dissertations, such studies are likely to be particularly subject to the “file drawer” problem. That is, post-hoc studies that fail to find expected positive effects are likely to be quietly abandoned, whereas studies supported by grants or produced as dissertations will almost always result in a report of some kind. The file drawer problem has been extensively described in research on meta-analyses

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education’s Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

and other quantitative syntheses (see, for example, Cooper, 1998), and it is a problem in all research reviews, but it is much more of a problem with post-hoc studies.

Despite all of these concerns, post-hoc studies were reluctantly included in the present review for one reason: without them, there would be no evidence at all concerning most of the commercial textbook series used by the vast majority of elementary schools. As long as the experimental and control groups were well matched at pretest on achievement and demographic variables, and met other inclusion requirements, we decided to include them, but readers should be very cautious in interpreting their findings. Prospective studies, in which experimental and control groups were designated in advance and outcomes are likely to be reported whatever they turn out to be, are always to be preferred to post-hoc studies, other factors being equal.

Another methodological limitation of almost all of the studies in this review is analysis of data at the individual student level. The treatments are invariably implemented at the school or classroom levels, and student scores within schools and classrooms cannot be considered independent. In clustered settings, individual-level analysis does not introduce bias, but it does greatly overstate statistical significance, and in studies involving a small number of schools or classes it can cause treatment effects to be confounded with school or classroom effects. In an extreme form, a study comparing, say, one school or teacher using Program A and one using Program B may have plenty of statistical power at the student level, but treatment effects cannot be separated from characteristics of the schools or teachers.

Several studies did randomly assign groups of students to treatments but nevertheless analyzed at the individual student level. The random assignment in such studies is beneficial, because it essentially eliminates selection bias. However, analysis at the student level, rather than the level of random assignment, still confounds treatment and school/classroom effects, as noted earlier. We call such studies "randomized quasi-experiments," and consider them more methodologically rigorous, all other things being equal, than matched studies, but less so than randomized studies in which analysis is at the level of random assignment.

Some of the qualifying studies, especially of instructional process programs, were quite small, involving a handful of schools or classes. Beyond the problem of confounding, small studies often allow the developers or experimenters to be closely involved in implementation, creating far more faithful and high-quality implementations than would be likely in more realistic circumstances. Unfortunately, many of the studies that used random assignment to treatments were very small, often with just one teacher or class per treatment. Also, the "file drawer" problem is heightened with small studies, which are likely to be published or otherwise reported only if their results are positive (see, Cooper, 1998).

Another methodological problem inherent to research on alternative mathematics curricula relates to outcome measures. In a recent criticism of the What Works Clearinghouse (WWC), Schoenfeld (2006) expressed concern that because most studies of mathematics curricula use standardized tests or state accountability tests focused more on traditional skills than on concepts and problem solving, there is a serious risk of "false negative" errors, which is to say that studies might miss true and meaningful effects on unmeasured outcomes

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

characteristic of innovative curricula. This is indeed a serious problem, and there is no solution to it. Measuring content taught only in the experimental group risks “false positive” errors, just as use of standardized tests risks “false negatives.” In the present review, only outcome measures that assess content likely to have been covered by all groups are considered; measures inherent to the treatment are excluded. However, many curriculum studies include outcomes for subscales, such as computations, concepts and applications, and problem solving, and these outcomes are separately reported in this review. Therefore, if an innovative curriculum produces, for example, better outcomes on problem solving but no differences on computations, that might be taken as an indication that it is succeeding at least in its area of emphasis.

A total of 87 studies met the inclusion criteria. This review discusses the methodological strengths and limitations of each study, which should be taken into account in understanding the outcomes. No single study is perfect, however, and the discussion of design limitations should be taken in context as an attempt to qualify findings, not to validate or invalidate them.

Tables 1-3 list all the qualifying studies. Within sections on each program, studies that used random assignment (if any) are discussed first, then randomized quasi-experiments, then prospective matched studies, and finally post-hoc matched studies. Within these categories, studies with larger sample sizes are listed first.

Studies of Mathematics Curricula

Perhaps the most common approach to reform in mathematics involves adoption of reform-oriented textbooks, along with appropriate professional development. Programs that have been evaluated fall into three categories. One is programs developed under funding from the National Science Foundation (NSF), that emphasize a constructivist philosophy, with a strong emphasis on problem solving, manipulatives, and concept development, and a relative de-emphasis on algorithms. At the opposite extreme is *Saxon Math*, a back-to-the-basics curriculum that emphasizes building students’ confidence and skill in computations and word problems. Finally, there are traditional commercial textbook programs.

The reform-oriented programs supported by NSF, especially *Everyday Mathematics*, have been remarkably successful in making the transition to widespread commercial application. Sconiers, Isaacs, Higgins, McBride, & Kelso (2003) estimated that in 1999, 10% of all schools were using one of three programs that had been developed under NSF funding and then commercially published. That number is surely higher as of this writing. Yet experimental-control evaluations of these and other curricula that meet the most minimal standards of methodological quality are very few. Only five studies of the NSF programs met the inclusion standards, and all but one of these was a post-hoc matched comparison.

This section reviews the evidence on mathematics curricula. Overall, 13 studies met the inclusion criteria, of which only two used random assignment. Table 1 summarizes the methodological characteristics and outcomes of these studies.

==============

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education’s Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Insert Table 1 Here

===============

The ARC Center Tri-State Study

The largest study by far of the NSF-supported mathematics curricula for elementary schools was carried out by the ARC (Alternatives for Rebuilding Curricula) Center at the Consortium for Mathematics and Its Applications (COMAP) (Sconiers, Isaacs, Higgins, McBride, and Kelso, 2003). This study located schools that had been using one of three NSF-funded curricula, *Everyday Mathematics, Math Trailblazers,* and *Investigations in Number, Data, and Space.* According to Sconiers et al., these reform-oriented curricula were united by a common set of goals. The developers claim that their curricula:

- Build on children's experiences;
- Teach basic arithmetic as well as geometry, data analysis, measurement, probability, and concepts of algebra;
- Challenge students with engaging and meaningful applications;
- Connect topics within mathematics and with other subjects;
- Encourage students to solve problems in many ways;
- Balance skills with concepts and problem solving;
- Include a variety of instructional approaches;
- Help teachers extend their understanding of mathematics and teaching; and
- Provide a variety of assessment instruments and procedures.

Developed in the 1980's, all three of these curricula were commercially published in the 1990's and are widely used alternatives to traditional commercial textbooks.

The study design involved identifying schools that had used any of the three programs, according to publishers' records, and matching them with schools in the same state that had not used the programs. Schools were sent a survey asking how long they had used the program, among other things, and schools had to have been using their programs for at least two years to be eligible for the study. The outcome variables were state test scores from the three states involved, Illinois, Massachusetts, and Washington, in the grades they tested (3, 4, and/or 5). Across the three states, a total of 742 schools were identified as implementing one of the reform programs.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Matching was carried out separately in each state, using variables available in that state. In all states, however, the main matching variables were reading scores, low income or free/reduced lunch percent, and percent White.

The results showed consistent positive effects for the reform curricula, but the effect sizes were quite modest. Across all states and grades, effect sizes for math scale scores averaged +0.10, with a range from +0.07 (Washington, Grade 3), to +0.12 (Illinois, Grade 5). This is equivalent, according to the authors, to a percentile change of less than 4% (e.g., from the 50th to the 54th percentile). However, due to the very large samples involved (from 6,879 to 14,875 in each state/grade) and analysis at the individual level, all differences were highly significant ($p<.001$).

Effect sizes were somewhat higher in measurement (+0.14) and lower in probability and statistics (+0.03). Effects were not higher, as might have been expected, on problem solving, the main focus of the three reform curricula. In Washington, which reports separate problem-solving scores, effect sizes for problem solving averaged a trivial +0.04 in third grade and +0.09 in fourth grade.

Overall effect sizes were similar for all ethnic groups except Hispanics, for whom differences averaged only +0.02, which was not significant. Differences were also similar for schools that were high, middle, and low in socioeconomic status, and were nearly identical for boys and girls.

Surprisingly, Sconiers et al. did not show outcomes separately for each curriculum, but the developers later produced their own reports. Those are discussed later in this review.

Everyday Mathematics

Everyday Mathematics, currently published by the Wright Group/McGraw-Hill, is the oldest, most comprehensively researched and most widely used of the NSF-supported reform curricula. It was developed beginning in the mid 1980s at the University of Chicago and published grade by grade over a ten-year period by the Everyday Learning Corporation.

Everyday Mathematics is based on the principles common to the NSF-supported reform models, described earlier.

Four studies of *Everyday Mathematics* met the standards of this review. Only one, a tiny study (38 students) among low-performing children, used a prospective matched design (Woodward & Baxter, 1997). It found no differences between *Everyday Mathematics* and control students (ES= -0.25). The other studies all used post-hoc matched designs. One, reported by SRA/McGraw-Hill (2003), separately analyzed data from the ARC study (Sconiers et al., 2003). Since most of the schools in that study used *Everyday Mathematics*, it is not surprising that the results were very similar to that of the larger study. The overall effect size was +0.12. Slight differences in outcomes by subscale, race, and socioeconomic status, are summarized in Table 1.

Riordan & Noyce (2001) carried out a post-hoc study of all Massachusetts schools that had used *Everyday Mathematics* for at least two years, in comparison to matched schools. Effect sizes were modest for schools that had used the program for 2-3 years (ES= +0.15), but among 19 schools that had used the program for four or more years, effect sizes were +0.35. This difference may be due to a cumulative effect of the program, but it may also be due to a selection artifact, as schools that began when these 19 schools began but did not experience success were likely to have dropped the program. Finally, Waite (2000) did a post-hoc comparison of schools using *Everyday Mathematics* in an urban Texas district. Out of 17 schools that began the program, only 7 continued to use it at the end of the year, and one of these refused to participate in the study. The possibility of selection bias is therefore substantial; whatever enabled these six schools to stick with the program may have made them more effective schools with any program. The effect size for this study was +0.26.

Across the studies there was no pattern of differential effects by measure, a surprising finding given the focus on concepts and problem solving. There were also no differences by ethnicity, except that in the SRA (2003) and Waite (2000) studies, effects for Hispanic students were near zero.

Math Trailblazers

Math Trailblazers (Carter, Beissinger, Cirulis, Gartzman, Kelso, & Wagreich, 2003) is an NSF-supported program for grades 1-5 that integrates mathematics with science, engaging children in laboratory investigations to experiment with mathematical ideas. For example, a second-grade unit involves filling containers with marshmallows from a modified egg carton that can contain ten groups of ten marshmallows (Carter et al., 2003).

The only qualifying study of *Math Trailblazers* is an analysis of data from the ARC study from Washington State (Kelso, Booton, & Majumdar, 2003). Separating out third- and fourth-grade scores for students who experienced *Math Trailblazers* and their matched controls, there were statistically significant differences favoring *Math Trailblazers* on total ITBS math scores, but in effect sizes terms they were extremely small, averaging +0.06 overall. There were no differences on computations (ES=0.00), and differences on concepts and estimation (ES=+0.09) and problem solving (ES=+0.07) were statistically significant but small. On the Washington Assessment of Student Learning (WASL), differences were also statistically significant but also very small, averaging ES=+0.05.

Saxon Math

Saxon Math is a K-12 curriculum that emphasizes incremental and explicit instruction, constant review, and frequent assessment. It is typically described as the antithesis of constructivist approaches, such as those supported by NSF, as it emphasizes mastery of mathematical algorithms, drill, and review.

The only qualifying study of *Saxon Math* was carried out by a third-party research company engaged by the publisher (Resendez & Azin, 2005). They identified 170 schools throughout the state of Georgia, of which 124 served students in grades 1-5. They then identified a similar number of control schools, matched on socioeconomic status, race, percentages of students with disabilities, and other variables. They obtained from state records school-level

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

means on Georgia's Criterion-Referenced Competency Test (CRCT) in mathematics, focusing on gains in successive cohorts of students from 2000 to 2005. At the elementary level, there were no differences in CRCT gains (ES=+0.02, n.s.). Although there were some small differences on some subscales at some elementary grade levels, there were no overall score differences for any ethnic group or for students with disabilities.

Scott Foresman-Addison Wesley

Scott Foresman-Addison Wesley (SFAW) Mathematics, a traditional K-6 mathematics text, was evaluated in two studies by the same independent contractor. Resendez & Azin (2006) carried out a randomized evaluation of *Scott Foresman Addison-Wesley Mathematics* in grades 3 and 5 in four schools in Ohio and New Jersey. Teachers (N=39) within these schools were randomly assigned to use SFAW or control (alternative basal texts) for a year. Students were pre- and posttested on Terra Nova Math Total (primarily word problems) and Math Computation in a one-year experiment. Multilevel modeling with students nested within teachers was used for the analyses. Results indicated no differences, adjusting for pretests. Effect sizes were -0.07 for Math Total and +0.05 for Math Computation, for an overall effect size of -0.01.

In a very similar study, Resendez & Sridharan (2005) randomly assigned teachers to use SFAW or control curricula, and then appropriately analyzed the data using hierarchical linear modeling (HLM), making this a true randomized evaluation. The study involved 35 teachers (18 SFAW, 17 control) in six schools. Random assignment was done within schools, among second and fourth grade teachers separately. HLM analyses found no significant differences in outcomes. Student-level effect sizes adjusted for pretests found slightly positive effect sizes for Math Total (ES= +0.10) but a negative effect size for Math Computations (ES= -0.21), for a mean effect size of -0.06.

Houghton Mifflin Mathematics

Houghton Mifflin Mathematics is a traditional textbook program. The publisher contracted with EDSTAR, an independent consulting company, to evaluate the program. Using a post-hoc design, EDSTAR identified eight California districts that had adopted *Houghton Mifflin Mathematics* and then identified control districts using other programs based on prior math scores, SES, ethnicity, and district size. State STAR test scores were obtained for 2001 (pre) and 2002 (post), for grades 2-5. Differences statistically favored *Houghton Mifflin*, but the overall effect size was only +0.14.

Growing with Mathematics

Growing with Mathematics is a standard mathematics program published by Wright Group/McGraw Hill. The publisher commissioned an evaluation of *Growing with Mathematics* by the University of Oklahoma (2004). A one-year post-hoc matched evaluation compared six schools using the program to five control schools. The schools were in Hawaii, Iowa, Oklahoma, and New Jersey. Students were in grades K-5. On the Terra Nova Comprehension scale, the *Growing with Mathematics* group gained more than the control group (ES = +0.20). On the Computations scale, the effect size was +0.23, for an overall effect size of +0.22.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

<u>*Excel Math*</u>

Excel Math is a K-6 mathematics curriculum that focuses on problem solving, integrated lessons, and development of thinking skills. Mahoney (1990) evaluated *Excel Math* in a post-hoc matched study in second and fourth grade classes in six California schools. Students were pre- and posttested on the Stanford Achievement Test. There were significant differences controlling for pretests favoring *Excel Math* in second grade (ES= +0.27) but not in fourth grade (ES= -0.02). The mean was ES= +0.13.

<u>*MathSteps*</u>

MathSteps was designed as a supplemental program focusing on computational skills, to remedy a perceived deficiency in constructivist mathematics curricula. It is specifically designed to help children perform successfully on norm-referenced standardized tests, providing structured activities for skillbuilding.

The publisher of *MathSteps*, Houghton Mifflin, contracted with Abt Associates to evaluate *MathSteps*. Chase, Johnston, Delameter, Moore, & Golding (2000) carried out an evaluation in five California districts. Despite its design as a supplementary program, Chase et al. discovered that more than two-thirds of teachers reported using *MathSteps* as their main text.

Chase et al. (2000) identified 21 elementary schools across the five districts and then identified 18 matched control schools in other California districts. They obtained SAT-9 data from state records for grades 3-5 in 1999 and 2000 to evaluate growth in *MathSteps* and control schools. *MathSteps* and control schools were matched on pretest scores, ethnicity, English language learners, and free lunch. Analysis of gain scores found that gains were not significantly different in *MathSteps* and control schools (overall ES=+0.03).

<u>*Knowing Mathematics*</u>

Knowing Mathematics is an intervention program published by Houghton Mifflin for students whose math skills are two or more years below grade level. An evaluation of *Knowing Mathematics* was carried out by the publisher (Houghton Mifflin, n.d.) in grades 4-6 in four schools in Lincoln, Nebraska. Students who participated in the program were compared to students in other schools matched on demographics and prior math achievement. All students were tested in spring, 2002 and spring, 2003 on the district's MAT/8 tests, although the intervention took place for only 12-14 weeks within the school year. Two of the study schools used *Knowing Mathematics* as an after school program and two used it as a pullout program during the school day.

Knowing Mathematics (N=21) and control students (N=18) were well matched at pretest. On posttests, subtracting out small pretest differences, *Knowing Mathematics* students scored somewhat better than controls on Total Mathematics (ES=+0.10), Computations (ES=+0.20), and Problem Solving (ES=+0.14). Because of the small sample sizes, none of these differences were statistically significant, however.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Conclusions: Mathematics Curricula

With a few exceptions, the studies comparing alternative mathematics curricula are of marginal methodological quality. Ten of the 13 qualifying studies used post-hoc matched designs in which control schools, classes, or students were matched with experimental groups after outcomes were known. Even though such studies are likely to overstate program outcomes, the outcomes reported in these studies are modest. The median effect size was only +0.10. The enormous ARC study found an average effect size of only +0.10 for the three most widely used of the NSF-supported mathematics curricula, taken together. Riordan & Noyce (2001), in a post-hoc study of *Everyday Mathematics*, did find substantial positive effects (ES=+0.34) in comparison to controls for schools that had used the program for 4-6 years, but effects for schools that used the program for 2-3 years were much smaller (ES=+0.15). This finding may suggest that schools need to implement this program for 4-6 years to see a meaningful benefit, but the difference in outcomes may just be a selection artifact, due to the fact that schools that were not succeeding may have dropped the program before their fourth year. The evidence for impacts of all of the curricula on standardized tests is thin. The median effect size across five studies of the NSF-supported curricula is only +0.12, very similar to the findings of the ARC study.

The reform-oriented math curricula may have positive effects on outcomes not assessed by standardized tests, as suggested by Schoenfeld (2006). However, the results on standardized and state accountability measures do not suggest differentially strong impacts on outcomes such as problem solving or concepts and applications that one might expect, as these are the focus of the NSF curricula and other reform curricula.

Evidence supporting *Saxon Math*, the very traditional, algorithmically-focused curriculum that is the polar opposite of the NSF-supported models, was lacking. The one methodologically adequate study evaluating the program, by Resendez & Azin (2005), found no differences on Georgia state tests between elementary students who experienced *Saxon Math* and those who used other texts.

More research is needed on all of these programs, but the evidence to date suggests a surprising conclusion that despite all the heated debates about the content of mathematics, there is limited high-quality evidence supporting differential effects of different math curricula.

Computer-Assisted Instruction

A longstanding approach to improving the mathematics performance of elementary students is computer-assisted instruction, or CAI. Over the years, CAI strategies have evolved from limited drill-and-practice programs to sophisticated integrated learning systems (ILS), which combine computerized placement and instruction. Typically, CAI materials have been used as supplements to classroom instruction, and are often used only a few times a week. Some of the studies of CAI in math have involved only 30 minutes per week. What CAI primarily adds is the ability to identify children’s strengths and weaknesses and then give them self-instructional exercises designed to fill in gaps. In a hierarchical subject like mathematics, especially computations, this may be of particular importance.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education’s Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

A closely related strategy, computer-managed learning systems (CMLS), is also reviewed in this section as a separate subcategory.

As noted earlier, CAI is one of the few categories of elementary mathematics interventions that has been reviewed extensively. Most recently, for example, Kulik (2003) reviewed research on uses of computer-assisted instruction in reading and math, and concluded that studies supported the effectiveness of CAI for math but not for reading. Murphy et al. (2002) concluded that CAI was effective in both subjects, but with much larger effects in math than in reading.

The following sections discuss research on several approaches to CAI in elementary mathematics. Many of these involved earlier versions of CAI that no longer exist, but it is still useful to be aware of the earlier evidence, as many of the highest-quality studies were done in the 1980's and early '90s. Overall, 38 studies of CAI met the inclusion criteria, and 15 of these used randomized or randomized quasi-experimental designs. In all cases, control groups used non-technology approaches, such as traditional textbooks. Table 2 summarizes the study characteristics and outcomes for CAI studies that met the inclusion criteria.

=================

Table 2 Here

=================

<u>*Jostens/Compass Learning*</u>

Jostens Learning System, now *Compass Learning*, provides students with a developmentally sequenced series of lessons in the basic curriculum areas of math, reading, and language arts. The curriculum is designed to correlate with basal texts, standardized tests, and district/classroom objectives. It emphasizes higher-order competencies, such as mathematical estimation and problem solving, as well as computations.

A total of 7 studies of *Jostens* met the inclusion criteria, more than any other CAI program. However, all of these evaluated early forms of *Jostens*, mostly from the 1980s. The current *Compass Learning* software has not been evaluated.

Two high-quality randomized studies evaluated the *Jostens* ILS. Alifrangis (1991) randomly assigned children in grades 4-6 to classes, and then their classes were randomly assigned to use *Jostens* software either in reading or in math. Those in the reading group therefore served as the control group in the math evaluation. On CTBS posttests, controlling for pretests, there were no differences (ES= -0.08). In a very similar study, Becker (1994) compared children in grades 2 to 5 randomly assigned to use *Jostens* ILS in reading or math. On CAT posttests, controlling for pretests, effect sizes were not significant, and averaged +0.05.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Two prospective matched studies evaluated *Jostens*. Zollman, Oldham, & Wyrick (1989) found significant positive effects on the MAT 6 (ES= +0.21) for students in grades 4-6. Hunter (1994) found larger positive effects with grades 2-5 (ES= +0.40).

Three post-hoc comparisons found mixed effects of *Jostens*. The largest, a statewide study by Estep, McInerney, Vockell, & Kosmoski (1999/2000), found no differences (mean ES= -0.01). A smaller study by Spencer (1999) found positive effects on CAT (ES= +0.39), and Clariana (1994) found positive effects on CTBS (ES= +0.66).

Across the 7 qualifying studies, the median effect size for *Jostens* was +0.21. However, the two highest-quality studies, randomized experiments by Alifrangis (1991) and Becker (1994), found no differences, making overall conclusions less clear.

CCC/SuccessMaker

SuccessMaker is an integrated learning system developed by the Computer Curriculum Corporation (CCC) and currently marketed by Pearson Digital Learning. It incorporates curriculum, management, and assessment in one package. *SuccessMaker*'s *Math Concepts and Skills*, designed for the K–8 mathematics curriculum, is intended for users who can use the courseware in 10- to 20-minute sessions three to five days per week to supplement classroom instruction.

Evaluations of CCC software all involved early forms of the program, mostly from the 1980s. The current version distributed by Pearson has not been evaluated.

Ragosta (1983) reported a randomized, three-year longitudinal evaluation (beginning in 1976) of an early form of CCC drill and practice software in four Los Angeles elementary schools. Within the four schools, students in Grades 2, 4, and 6 initially received CAI. They were randomly assigned to experience math only, reading/language arts only, or math and reading/language arts. In each case, students received two 10-minute sessions of CAI daily. Students were pre and post tested on the CTBS. The experimental design was complex, but combining across grades, there were strong positive effects of the CCC math strand on CTBS Computations (ES = +0.72), but nonsignificant differences on Concepts (ES = +0.09) and Applications (ES = +0.26). Averaging the three scales, the effect size was +0.36.

Another randomized evaluation of early CCC software was carried out by Hotard & Cortez (1983) in two high-poverty schools in Louisiana. Students in grades 3-6 were randomly assigned to CAI or non-CAI conditions. Those in the CAI classes worked with CCC software ten minutes each day, in addition to their math lab time, for six months. Posttests, controlling for pretests, were significantly higher for the CAI group (ES=+0.19, $p<.01$).

A dissertation by Manuel (1987) involved 165 children in grades 3-6 in three schools in Omaha, Nebraska. The 12-week study compared students who experienced 10 minutes per day of CCC or of a variety of math software on Apple computers. A control group did not use CAI. Students within the three study schools were randomly assigned to the three conditions, stratifying on ability and pretest measures. Analyses of gain scores indicated no significant differences among conditions ES=+0.07, n.s.).

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Mintz (2000) evaluated *SuccessMaker* in grades 4-5 in Etowah County, Alabama. Schools using *SuccessMaker* as a supplement to *Saxon Math* were compared to schools using *Saxon Math* without CAI. On SAT-9 posttests adjusted for pretests, there were no differences in grade 4 (ES=+0.08) or grade 5 (ES= -0.20), for a mean effect size of -0.06.

A post-hoc matched study by Laub (1995) compared students' fall-to-fall gains on the Stanford Achievement Test in the two years before *SuccessMaker* was adopted to gains in 1993-94. Gains were significantly larger in the year *SuccessMaker* was introduced (ES=+0.27, p<.001).

Across the five qualifying studies of *SuccessMaker*, the median effect size was +0.19, but the studies with positive effects evaluated very early forms of the program unlike those that exist today.

Classworks

Classworks, designed by Curriculum Advantage, is a comprehensive computer learning system. It contains over 1,000 units of instruction, drawn from over 100 software titles. *Classworks* provides comprehensive curriculum materials, as well as the tools that let teachers and administrators manage, assess, and individualize their students' learning process.

A tiny but randomized experiment by Patterson (2005) compared 30 third graders randomly assigned to be taught for 14 weeks using either *Classworks* as a supplement to *Saxon Math* or *Saxon Math* alone. *Classworks* was used only one hour per week. The control class used *Classworks* language arts, but not math. On the SAT-9, the effect size was +0.85.

Whitaker (2005) carried out a small post-hoc evaluation of *Classworks Gold* in grades 4-5 in two rural Tennessee schools. The study compared math and reading gains on TCAP in two Tennessee schools. Posttest differences did not approach statistical significance (ES=+0.21).

Lightspan

Lightspan, currently distributed by Plato Learning, Inc. under the Plato name, is a K-6 curriculum-based technology program designed to be used in both schools and homes. The home-involvement program incorporates parent training, loaning hardware (Sony's PlayStation®) to parents, and sending home *Lightspan Adventures* (learning games) with students.

A two-year study by Birch (2002) evaluated *Lightspan* in two Wilmington, Delaware elementary schools. In 1997-98, second and fifth graders in the experimental school began to use the program, and were compared to matched control schools. Students were pre- and posttested on the Stanford Achievement Test (SAT). Using analyses of covariance, the *Lightspan* students significantly outscored the control students at the end of Year 1 (ES=+0.53, p<.01) and, to a lesser extent, at the end of the second year (ES=+0.28, p<.01).

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Other Computer-Assisted Instruction

Several qualifying studies evaluated forms of CAI that are no longer disseminated. However, the findings of these studies are still useful in understanding effects of CAI. Their characteristics and results are summarized in Table 2.

Computer-Managed Learning Systems

A set of strategies related to CAI, computer-managed learning systems (CMLS) are programs that use computers to identify students' needs, assign them appropriate exercises, assess outcomes, and communicate to teachers information about student performance. There is only one CMLS program, *Accelerated Math*, that has been evaluated in qualifying studies.

Accelerated Math

Accelerated Math (Renaissance Learning, 1998a) is a supplementary approach to mathematics instruction that uses computers to assess children's levels of performance, and then generates individualized assignments appropriate to their needs. Students scan completed assignments into the computer, which gives teachers regular diagnostic reports that they are expected to use to develop targeted interventions. The *Accelerated Math* curriculum focuses on foundational skills, especially computations, and is intended for use along with other traditional or reform-oriented math programs.

Accelerated Math has been evaluated in several experiments of good quality, published in peer-reviewed journals, but they share a characteristic that makes some of their findings ineligible for inclusion in this review. This is the use of a measure called *STAR Math* as the outcome variable. *STAR Math* (Renaissance Learning, 1998b) is a computer-adaptive test used in the *Accelerated Math* program itself to regularly assess and place students. It is closely aligned with the objectives and format of *Accelerated Math*, and students would have practiced the unusual format as part of their experience with the program. For this reason, studies are only included if they also used measures other than *STAR Math* that were fair measures of what both *Accelerated Math* and control students experienced.

Ysseldyke & Bolt (2006) carried out a large randomized quasi-experimental evaluation of *Accelerated Math*. Teachers of grades 2-5 in five schools in Texas, Alabama, South Carolina, and Florida were randomly assigned to *Accelerated Math* or control conditions in a one-year experiment. Results indicated no differences on Terra Nova mathematics controlling for pretests (ES=+0.03, n.s.).

A large post-hoc matched study in southern Mississippi (Ross & Nunnery, 2005) compared 23 elementary schools using the *School Renaissance* comprehensive reform model, which includes both *Accelerated Reader* and *Accelerated Math*, to 18 matched schools. The Mississippi Curriculum Test was used as a pre- and posttest. Analyses of covariance found no significant differences in math posttest scores in grades 3-5 (mean ES=+0.04).

Two prospective matched studies evaluated Accelerated Math in Minneapolis. One (Ysseldyke et al., 2003) found small differences (ES=+0.12) on the Northwest Achievement Test Level (NALT). Another (Spicuzza et al., 2001) found adjusted NALT scores to be higher on within-school comparisons (ES=+0.19) and district comparisons (ES=+0.14).

Johnson-Scott (2006) evaluated *Accelerated Math* in predominately African-American, 100% free lunch schools in rural Mississippi. Fifth graders using *Accelerated Math* to supplement a *McGraw-Hill* textbook were compared to students who used the same text without *Accelerated Math*. The Mississippi Curriculum Test (MCT) was used as a pre- and posttest. The control group scored somewhat higher at pretest, but the *Accelerated Math* group scored slightly higher at posttest, for an adjusted effect size of +0.23.

Across all five studies of *Accelerated Math*, the median effect size on independent measures was +0.11.

Conclusions: Computer-Assisted Instruction

In sheer numbers of studies, computer-assisted instruction is the most extensively studied of all approaches to math reform. A total of 38 qualifying studies were identified, of which 15 used random assignment. Most studies of CAI find positive effects, especially on measures of math computations. Across all studies from which effect sizes could be computed, these effects are meaningful in light of the fact that CAI is a supplemental approach, rarely occupying more than three 30-minute sessions weekly (and often alternating with CAI reading instruction). The median effect size was +0.19. This is larger than the median found for the curriculum studies (+0.10), and it is based on many more studies (38 vs. 13) and on many more randomized and randomized quasi-experimental studies (15 vs. 2). However, it is important to note that most of these studies are quite old, and usually evaluated programs that are no longer commercially available.

While outcomes of studies of CAI are highly variable, most studies do find positive effects, and none significantly favored a control group. While the largest number of studies has involved Jostens, there is not enough high quality evidence on particular CAI approaches to recommend any one over another, at least based on student outcomes on standardized tests.

In studies that break down their results by subscales, outcomes are usually stronger for computations than for concepts or problem solving. This is not surprising, as CAI is primarily used as a supplement to help children with computations skills. Because of the hierarchical nature of math computations, CAI has a special advantage in this area because of its ability to assess students and provide them with individualized practice on skills that they have the prerequisites to learn but have not yet learned.

Instructional Process Strategies

Many researchers and reformers have sought to improve children's mathematics achievement by giving teachers extensive professional development on the use of instructional process strategies, such as cooperative learning, classroom management, and motivation strategies (see Hill, 2004). Curriculum reforms and CAI also typically include professional

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

development, of course, but what primarily characterizes the strategies reviewed in this section is a focus on changing what teachers do with the curriculum they have, not changing the curriculum.

A total of 36 studies evaluated instructional process programs. Of these, 19 studies used randomized or randomized quasi-experimental designs. Table 3 summarizes characteristics and outcomes of these studies.

==================

Table 3 Here

==================

The programs in this section are highly diverse, so they are further divided into seven categories:

1. Cooperative Learning
2. Cooperative/Individualized Programs
3. Direct Instruction
4. Mastery Learning
5. Professional Development Focused on Math Content
6. Professional Development Focused on Classroom Management and Motivation
7. Supplemental Programs

Cooperative Learning

Cooperative learning refers to a set of strategies in which students work in pairs or small groups to help each other master academic content. Extensive research on cooperative learning in many subjects and grade levels has found that cooperative learning increases student learning if it provides the groups with a common goal that they can only achieve if all group members do well on independent assessments (Johnson & Johnson, 1998; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003; Slavin, 1995; Slavin, Hurley, & Chamberlain, 2001). That is, cooperative learning enhances achievement when it motivates children to teach each other, in a setting in which teammates know that their own success depends in part on the learning of their teammates (see Webb & Palincsar, 1996). The following sections summarize the findings of elementary studies of cooperative learning.

Classwide Peer Tutoring

Classwide Peer Tutoring, or *CWPT* (Greenwood, Delquadri, & Hall, 1989), is a cooperative learning approach in which children regularly work in pairs. They engage in

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

structured tutoring activities and frequently reverse roles. The pairs are grouped within two large teams in each classroom, and tutees earn points for their team by succeeding on their learning tasks. A winning team is determined each week, and receives recognition.

A remarkable four-year longitudinal study by Greenwood et al. (1989) evaluated *CWPT*. In it, six high-poverty schools in Kansas City, Kansas, were randomly assigned to *CWPT* or control conditions. Because analysis was at the student level, this was a randomized quasi-experiment. The children and teachers began in Grade 1 and continued through Grade 4. A total of 123 students began in the experimental and control schools in first grade and continued through fourth grade, about half of the initial group. Otis-Lennon IQ scores for the two groups in the longitudinal sample were nearly identical, and there were no significant differences in SES.

At posttest, analyses of covariance indicated significantly higher achievement for the *CWPT* group on the mathematics section of the Metropolitan Achievement Test (ES=+0.33, p<.02). A two-year followup, when children were in sixth grade, found that CWPT students maintained their advantage over the control students (ES=+0.23) (Greenwood & Terry, 1993).

Peer-Assisted Learning Strategies

In *Peer-Assisted Learning Strategies*, or *PALS* (Fuchs, Fuchs, & Karns, 2001; Fuchs, Fuchs, Phillips, Hamlett, & Karns, 1995; Fuchs, Fuchs, Yazdian, & Powell, 2002), children work in pairs to learn mathematical concepts with each other. Children alternate every 15 minutes as tutor and tutee, using specific strategies for correction procedures. *PALS* is used as a supplement to traditional textbook-based instruction approximately 30 minutes a day, three times a week.

Fuchs, Fuchs, Yazdian, & Powell (2002) evaluated PALS with first graders. Within schools in a southeastern city, 20 teachers were randomly assigned to *PALS* or non-*PALS* conditions. *PALS* was used for 16 weeks to supplement a basal program called *Math Advantage*. Non-*PALS* classes used *Math Advantage* only. Since data were analyzed at the student level, this was a randomized quasi-experiment.

Items from the Stanford Achievement Test were used as pre- and posttests. On gain scores, *PALS* students scored only modestly more than non-*PALS* students on the total test, a difference that is only marginally significant (ES= +0.11, p=.068).

In a similar randomized quasi-experiment with kindergartners, Fuchs, Fuchs, & Karns (2001) randomly assigned 20 kindergarten teachers within five schools in a southeastern city to use *PALS* or to continue using traditional methods. Two standardized measures were used. The *SESAT*, a mathematics readiness measure, was used as a pre- and posttest. The Stanford Achievement Test was used as a posttest only. The two groups were similar at pretest.

On SESAT posttests, the *PALS* children showed significantly greater gains in a time by condition ANOVA (ES=+0.24, p<.05). On the SAT, differences were not statistically significant, but effect sizes also averaged around +0.24.

Student Teams-Achievement Divisions (STAD) and Student Team Mastery Learning

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Mevarech (1985) carried out a study of cooperative learning, mastery learning, and a combination of the two in a middle class Israeli elementary school. Fifth graders were randomly assigned to four conditions. *Student Team Learning (STL)* was based on a program called *Student Teams-Achievement Divisions (STAD)* (Slavin, 1995). As in STAD, following teacher instruction, students worked in four-member teams to help each other master mathematics content. Team scores were based on the sum of students' individual test scores, and the highest-scoring teams received small rewards. Mastery learning (ML), discussed in more detail later in this review, involved giving students formative tests following instruction. Those who scored less than 80% received additional instruction until they could pass the test. *Student Team Mastery Learning (STML)* involved a combination of the two strategies, and there was a control group that received traditional instruction. The 15-week study focused on fractions, and a 48-item pre- and posttest assessed the content that was emphasized equally in all four classes. All groups were very similar at pretest.

A 2 x 2 analysis of covariance, controlling for pretests, evaluated outcomes of the treatments separately for computations and word problems. All three experimental groups received significantly and substantially higher scores than the control treatment. Effect sizes for *STML* in comparison to control were +0.36 for computations and +0.21 for word problems; +0.36 and +0.01 for *STL*; and +0.55 and +0.28 for *ML*.

In a 12-week experiment, Mevarech (1991) further explored cooperative learning, mastery learning, and combinations of the two. Intact Israeli third grade classes in a low-SES school were randomly assigned to the same four treatments as in the Mevarech (1985) study. Students in all three treatments substantially exceeded the control group in test scores, controlling for pretests. Effect sizes for the combined STML group averaged +0.55, for STL +0.60, and for mastery learning +1.08.

Glassman (1989) evaluated *Student Teams-Achievement Divisions* (STAD) in two diverse schools in Long Island, New York. Third, fourth, and fifth grade classes were randomly assigned within schools to use STAD in reading, writing, and math, or to continue using traditional strategies, in a six-month randomized quasi-experiment. ITBS was used as a pre- and posttest. In mathematics, there were no differences in adjusted outcomes (ES=+0.01).

In a study involving 10 schools in rural Indonesia, Suyanto (1998) evaluated *STAD* in grades 3-5. Fifteen *STAD* classes were compared to 15 controls, which used traditional whole-class methods, over a 4-month experiment. *STAD* and control classes were well matched on the Indonesian Elementary School Tests of Learning. At posttest, adjusted for pretests, students in the *STAD* classes scored significantly higher (ES=+0.40, p<.001).

Conclusions: Cooperative Learning

Across the 9 qualifying studies of various cooperative learning strategies, the median ES was +0.29. Eight of the nine studies used randomized or randomized quasi-experimental designs.

Cooperative/Individualized Programs

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Two programs, *TAI Math* and *Project CHILD*, are core classroom instructional models that combine extensive use of cooperative learning with strategies for continuously diagnosing students' strengths and weaknesses and giving them material appropriate to their needs.

Team Assisted Individualization (TAI)

Team Assisted Individualization, or *TAI* (Slavin, Leavey, & Madden, 1986), uses individualized instruction and cooperative learning as core strategies. Developed in the early 1980's, it is still disseminated by Charlesbridge Publishing under the name *Team Accelerated Instruction.*

In *TAI*, designed for grades 3-6, students work in 4-5 member, heterogeneous teams. They are initially tested and placed in an instructional sequence according to their current levels of performance. Teachers introduce concepts in groups of students drawn from the teams who are at the same performance level. Students then work through individualized materials with the help of their teammates, preparing for individualized assessments. Teams receive certificates and recognition based on the progress made by all members in passing these assessments.

The *TAI* materials focus primarily on computations and problem solving. Teachers used the *TAI* materials three weeks each month, and then spent the fourth week teaching objectives such as geometry and measurement using other materials in a whole-class (not individualized) format.

Several studies using randomized and matched designs have evaluated outcomes of TAI. Two randomized evaluations were carried out by Slavin & Karweit (1985). The larger of the two studies, referred to by Slavin & Karweit (1985) as Experiment 2, took place in and around a rural town in Western Maryland. In this study, 17 classes were randomly assigned to *TAI*, *MMP*, or to an untreated control group for a 16-week period. Using nested analyses of covariance similar to hierarchical linear modeling (HLM), *TAI* students scored significantly higher than *MMP* on *CTBS* Computations (ES=+0.39, p<.01), but not on Concepts and Applications (ES=+0.01, n.s.), for a mean effect size of +0.20. In comparison to the untreated control group, effects favoring *TAI* for CTBS Computations were substantial (ES=+0.67, p<.001), but there were no effects for Concepts and Applications (ES=+0.06, n.s.).

The study referred to by Slavin & Karweit (1985) as Experiment 1 took place in inner-city Wilmington, Delaware. Students in 10 grade 4-6 classes were randomly assigned to *TAI* or to a whole-class instructional program used with traditional texts called the *Missouri Mathematics Program*, or *MMP* (Good, Grouws, & Ebmeier, 1983). A form of *MMP* that used within-class grouping was also evaluated. *MMP* is discussed in detail earlier in this review.

The experiment took place over an 18-week period. Students were well matched at pretest, although the *TAI* group scored non-significantly higher on California Achievement Test pretests. At posttest, using random effects nested analyses of covariance with classroom as the unit of analysis, *TAI* classes scored significantly higher than MMP classes on the Comprehensive Test of Basic Skills (CTBS) Computations scale (ES=+0.76, p<.001) but not on Concepts and Applications (ES=0.00, n.s.).

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

In the largest evaluation of *TAI* (Slavin, Madden, & Leavey, 1984) took place in a middle class suburb of Baltimore. Fifty-nine volunteer teachers of grades 3-5 were non-randomly assigned to *TAI* or traditionally-taught control conditions over a 24-week period. *TAI* and control schools were well matched on CAT pretests. Nested analyses of covariance found significant positive effects on CTBS Computations (ES=+0.18, p<.045), but not on Concepts and Applications (ES=+0.10, p<.0.12). It is important to note that individual-level analyses found statistically significant differences on both subscales (p<.001 in both cases). Individual-level analyses for the students with special needs found significant positive effects of *TAI* on CTBS Computations (ES=+0.19, p<.015) and Concepts and Applications (ES=+0.23, p<.04). There were no treatment x disability interactions.

Stevens & Slavin (1995) carried out a study of *TAI* as part of a schoolwide reform model called the Cooperative Elementary School, which also used cooperative learning in other subjects. The study took place over a two-year period in a diverse Baltimore suburb. Twenty-one grade 2-6 classes in two schools were matched with 24 classes in three comparison schools matched on average student achievement, ethnicity, and SES. Using hierarchical linear modeling (HLM), posttest means after two years were found to be significantly higher in the *TAI* classes than in control classes on CAT Computations (ES=+0.29, p<.01), but there were no significant differences on Applications (ES=+0.10, n.s.). For students with special needs, outcomes were particularly positive for both Computations (ES=+0.59, p<.01) and Applications (ES=+0.35, p<.05). Effects for gifted students were also very positive for Computations (ES=+0.59, p<.01) but not Applications (ES=+0.19, n.s.).

Karper & Melnick (1993) carried out a small year-long study of *TAI* in grades 4-5 in an affluent suburb of Harrisburg, Pennsylvania. *TAI* and control classes as each grade level were well matched on an unnamed district standardized test. At posttest there were no differences on the same test (ES= -0.11, n.s.).

The two large randomized experiments, among a total of 5 studies, and the substantial positive effects on computations measures in 4 of the studies, make TAI particularly well supported for its computations effects (median ES=+0.29), but positive effects for concepts and applications were not demonstrated (median ES=+0.04).

Project CHILD

Project CHILD (Orr, 1991) is a school restructuring program that engages students in classrooms that emphasize cooperative learning, self-regulated behavior, active learning, and technology integration. Students work in multiage groups (K-2 and 3-5) with a team of three teachers. One teacher on each team is the math specialist, so students have three years with one teacher in grades K-2 and one in grades 3-5. Students are given units of work in each subject appropriate to their developmental needs. A variety of CAI content is extensively used in each unit. Students in Project CHILD use cooperative learning, and keep records of their progress on individual “passports.” Children in each class rotate through learning stations, one of which has 3-6 computers.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education’s Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

The evaluation took place in two Florida schools. In both schools, Project CHILD was only used with a subset of children in each grade, and the evaluation compared children in Project CHILD to a matched subsample.

The Project CHILD and control students were well matched on unspecified standardized test scores across grades 3-5. Analyses of covariance showed significantly greater gains in math, controlling for pretests, for Project CHILD students (ES=+0.69, p<.001).

Direct Instruction

Direct Instruction is an approach to instruction that emphasizes a structured, step-by-step approach focusing on the "big ideas" of mathematics. Direct Instruction mathematics programs include *Connecting Math Concepts, DISTAR Arithmetic,* and a variation called *User-Friendly Direct Instruction.*

Connecting Math Concepts

Connecting Math Concepts, or *CMC* (Engelmann, Carnine, Kelly, & Engelmann, 1988-1994) is a mathematics curriculum developed by the authors of Direct Instruction at the University of Oregon. It is based on the concepts of *DISTAR Arithmetic*, which was part of the Direct Instruction treatment found in the classic Project Follow Through to substantially increase the reading, arithmetic, and language achievement of disadvantaged children in grades K-3 (Adams & Engelmann, 1996). *CMC* has six guiding principles of effective instruction: 1) key concepts, "big ideas", are taught that have broad applicability; 2) prerequisite skills are introduced before complex learning; 3) explicit instruction, with specific strategies and rules, is used to teach concepts, 4) guided practice is given to the students in the beginning stages of learning and phased out as students become more competent; 5) each new strategy is woven with other strategies in order to clearly connect different aspects of knowledge; and 6) cumulative review is provided. Teachers follow a detailed manual that gives them specific wording and error correction procedures to use in all lessons.

Snider & Crawford (1996) carried out a small randomized evaluation of *CMC* in two classes. Fourth graders were randomly assigned to a class using *CMC* or one using a Scott Foresman text for a year-long study. The main outcome measure was the National Achievement Test (NAT) in mathematics. On fall pretests there were few differences. At posttest, *CMC* students scored significantly and substantially better on computations (ES= +0.72, p< .001), but not on concepts and problem-solving (ES= +0.01, n.s.) or total score (ES=+0.26, n.s.).

Another evaluation of *Connecting Math Concepts* was done by Tarver & Jung (1995) in a Midwestern suburban elementary school. There were 119 first grade students assigned to one of five classes—one experimental CMC class and four control classes using a discovery-oriented mathematics approach. At posttest, the *CMC* students scored significantly higher than comparison students, controlling for pretests, on Computations (ES= +2.13, p<.05), Concepts and Applications (ES=+0.68, p<.05), and Total Math (ES=+1.33, p<.05).

Crawford & Snider (2000) evaluated *CMC* in a year-long study in which students in one teacher's fourth grade class were compared to those in the same teacher's class the previous year, when she was using a Scott Foresman textbook. The two classes were nearly identical on the National Achievement Test (NAT) at pretest, but at posttest the *CMC* students scored higher on NAT Computations (ES=+0.33), Concepts and Problem Solving (ES=+0.39), and Total (ES=+0.41). Because of the very small sample sizes and confounding of treatment with teacher effects, all three of the *CMC* studies should be interpreted with caution.

User-Friendly Direct Instruction

Grossen & Ewing (1994) evaluated an adaptation of *Direct Instruction* that they called *User-Friendly Direct Instruction*. It used a teacher-directed, structured approach built around a series of videodisc lessons called *Systems Impact*. Fifth graders (N=45) in a Rocky Mountain elementary school were randomly assigned to *User-Friendly Direct Instruction* or to use a standard *Scott Foresman* textbook over two school years. Four independent posttest measures adjusted for pretests favored the DI groups: Woodcock-Johnson Applications (ES=+0.50), ITBS Concepts (ES=+0.44), ITBS Problem Solving (ES=+0.17), and ITBS Operations (ES=+0.97). Because of the small sample size, only the Operations differences were statistically significant, but the overall average effect size was +0.52. Again, because of the small sample size and confounding with teacher and class effects, these results should be interpreted with caution.

Mastery Learning

Mastery learning (Block & Anderson, 1976) is an instructional strategy in which students are taught a given topic, usually for about two weeks, and then given a formative test. Students who do not pass the test at a predetermined level, typically 80% correct, are given corrective instruction intended to remediate their learning deficits. Those who do pass work on enrichment activities.

Only a few studies of mastery learning in elementary mathematics met the standards of this review. The evidence supporting the effectiveness of mastery learning comes from two Israeli studies described earlier (see *Student Teams-Achievement Divisions* and *Student Team Mastery Learning*, above) by Mevarech (1985, 1991). Both studies evaluated mastery learning, cooperative learning, and a combination of the two. The Mevarech (1985) study involved fifth graders randomly assigned to the three treatments or to a control group for 15 weeks. In comparison to controls, effect sizes for mastery learning averaged +0.55 for computations and +0.28 for word problems. In the 12-week randomized quasi-experiment by Mevarech (1991), which took place in Israeli third grades, the effect size favoring mastery learning was +1.08. However, there was only one teacher per treatment, so teacher effects were confounded with treatment effects.

Other studies of mastery learning have found far less positive outcomes. Monger (1989) compared matched schools in two suburban Oklahoma districts. The study involved 70 students in grade 2 and 70 in grade 5, a one-in-four sample selected systematically to represent approximately 280 students at each grade level. Students were fairly well matched on Metropolitan Achievement Tests (MAT6). On MAT6 posttests, there were no significant differences at either grade level. On MAT6 Total Mathematics, non-significant differences favored the control group in grade 2 (ES= -0.09) and 5 (ES= -0.27), for an average of -0.18.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Anderson, Scott, & Hutlock (1976) compared students in grades 1-6 in one mastery learning and one control school in near Cleveland, Ohio. Students were matched on the Metropolitan Readiness Test in grades 1-3, or the Otis-Lennon Intelligence Test in grades 4-6. After a year, students were tested on the California Achievement Test mathematics scales. There were no significant differences in outcomes. Mastery learning students gained somewhat more than controls on Computations (ES=+0.17) and Problem Solving (ES=+0.07), but not Concepts (ES= -0.12).

Cox (1986) evaluated a form of mastery learning in a school in Southwestern Missouri. Fifth graders were assigned by alternating down a ranked list to mastery learning or control conditions . The study took place over a 6-month period. Students were pre- and posttested on ITBS. Posttests adjusted for pretests showed an effect size of +0.22.

Professional Development Focused on Mathematics Content

Two programs, *Cognitively Guided Instruction* and *Dynamic Pedagogy*, provide teachers with extensive professional development focused on how children learn mathematics and how to help them build on their intuitive knowledge. Both can be used with any textbook or curriculum.

Cognitively Guided Instruction

Cognitively Guided Instruction, or CGI (Carpenter, Peterson, Chiang, & Loef, 1989; Carpenter, Fennema, Franke, Empson & Levi, 1999) is an approach to mathematics reform that uses extensive professional development to prepare elementary teachers to teach mathematics for understanding by building on the intuitive knowledge of mathematics and problem solving strategies that children bring to instruction.

Carpenter et al. (1989) carried out a year-long randomized evaluation of *CGI* in schools in and around Madison, Wisconsin. Schools were randomly assigned to *CGI* or control conditions, although teachers (20 *CGI*, 20 control) were the unit of analysis, making this a randomized quasi-experiment. First graders were pre- and posttested on the Iowa Test of Basic Skills (ITBS) supplemented by additional problem-solving items. On ITBS Computations, there were no significant differences at posttest, although it is highly likely that student-level analyses would have been significant (ES=+0.24). The same is true of ITBS problem solving: Differences approached statistical significance at the teacher level, but almost surely would have been significant but small at the student level.

Dynamic Pedagogy

Dynamic Pedagogy is a professional development model for mathematics in which teachers learn to prepare and deliver lessons appropriate to students' current knowledge, misconceptions, and past errors. Teachers establish clear objectives for each lesson and follow a lesson structure emphasizing use of manipulatives and movement from initiation to development to closure. Lessons focus on connecting new concepts to prior knowledge, multiple representations of math processes, correcting misconceptions, and frequently assessing student progress. Teachers meet regularly to receive feedback on their lessons from project staff.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

A year-long evaluation of *Dynamic Pedagogy* was carried out by Armour–Thomas, Walker, Dixon-Roman, Mejia, & Gordon (2006) in two majority African-American K-3 elementary schools. A total of 60 third graders were matched with non-participating students in the same two schools. Terra Nova Math posttests, controlling for pretests, significantly favored the *Dynamic Pedagogy* group (ES = +0.51).

Professional Development Focused on Classroom Management and Motivation

Two programs, the *Missouri Mathematics Project* and *Consistency Management & Cooperative Discipline* focus primarily on improving teachers' abilities to use effective instructional and management techniques, to make effective use of time, and to enhance student motivation.

Missouri Mathematics Project

The *Missouri Mathematics Project*, or *MMP* (Good, Grouws, & Ebmeier, 1983) is a program designed to help teachers effectively use practices that had been identified from earlier correlational research to be characteristic of teachers whose students made outstanding gains in achievement (Good & Grouws, 1979). The intervention focuses on teaching teachers to engage in active teaching with lively explanations, and a focus on meaning, moderate amounts of well-managed seatwork, daily review with mental mathematics exercises, frequent assessments, and a rapid pace of instruction.

As part of a larger evaluation of TAI, Slavin & Karweit (1985, Experiment 2) evaluated two variations of *MMP*. One involved whole-class instruction (*MMP*-whole class), and one used within-class grouping, with teachers alternating between two performance groups (*MMP*-two group). The study took place over a 16-week period in spring, 1983.

Seventeen teachers of grades 3-5 in and around a small city in Western Maryland were randomly assigned to *MMP*-whole class, *MMP*-two group, or control (traditional whole-class instruction). Their 366 students were posttested on CTBS, and district-administered CAT scores were used as pretests. Nested analyses of covariance, similar to HLM, were used to analyze differences among treatments. Students in *MMP*-whole class scored better than those in the control group, controlling for pretests (ES=+0.29, $p<.05$), and those in the *MMP*-two group treatment scored significantly better than MMP-whole class (ES=+0.55, $p<.001$) and substantially better than control (ES=+0.84, $p<.001$).

Good & Grouws (1979) evaluated *MMP* in schools in Tulsa, Oklahoma. The 27 schools randomly assigned to *MMP* or control conditions, with 40 fourth-grade teachers using *MMP* or control methods. All schools used the same textbooks and were similar in SES. This randomized quasi-experiment continued for a total of three months. Students were pre- and posttested on SRA-Mathematics. At pretest, the control group scored somewhat higher on SRA, but at posttest the *MMP* students scored significantly higher. Only teacher (not student) means and standard deviations were reported, but effect sizes could be estimated at ES=+0.33.

Ebmeier & Good (1979) evaluated the *Missouri Mathematics Program* with 39 fourth-grade math teachers in a large southwestern school district. These teachers taught 68 sections of

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

math across 28 mostly high-poverty schools. Students were pre- and posttested on the SRA Achievement Test. Analyses of covariance controlling for pretests found substantial positive effects of *MMP* (ES=+0.42, p<.01).

Across the three studies, all of which used random or randomized quasi-experimental designs, the median effect size was +0.42.

Consistency Management & Cooperative Discipline

Consistency Management & Cooperative Discipline, or *CMCD* (Freiberg, 1996, 1999) is a preventive approach to classroom management that emphasizes shared student and teacher responsibility for learning. It trains teachers in strategies for engaging students in setting and adhering to classroom rules, giving students helping roles within the classroom (such as taking attendance and passing out papers), involving parents, and using strategies for calling on students that ensure that all will have opportunities to respond.

Freiberg, Prokosch, Treiser, & Stein (1990) evaluated CMCD in five Houston elementary schools. The schools were 90% African American, and 72% of students qualified for free or reduced-price lunch. In a post-hoc matched comparison, five similar schools were identified, matched on prior test scores and demographics. District-administered standardized tests were followed from 1986 (pre) to 1988 (post).

Students (N=364) of 28 grade 2-5 teachers who had received full CMCD training and had remained in their schools from 1986 to 1988 were compared to a randomly selected group of students in the control schools (N=335). The groups were well matched on pretest scores and demographics. Posttests adjusted for pretests were significantly higher for the CMCD students on the MAT6 for grades 2-5 (ES=+0.29). On the Texas Education Assessment of Minimal Skills (TEAMS), students in grades 3 and 5 scored substantially better in CMCD schools than in control schools (ES=+0.51).

Freiberg, Connell, & Lorentz (2001) carried out a study in which *CMCD* was integrated with a constructivist mathematics curriculum called Move-it Math, used within a larger schoolwide reform program called Project GRAD. In 1994-95, three Project GRAD schools in inner-city Houston adopted *CMCD* along with Move-it Math, while four schools in the same middle school feeder system used the same math programs but without *CMCD*. Most students were Latino, and the two sets of schools were similar in demographic characteristics and pretest scores.

In a post-hoc study, Freiberg at al. (2001) compared Texas Assessment of Academic Skills (TAAS) mathematics scores for *CMCD* and non-*CMCD* students before and one year after implementation. Data were combined for all students in grades 4-6. Results indicated significantly greater gains for the *CMCD* students (ES=+0.33, p<.001).

A Newark study by Opuni (2006) compared schools that used CMCD to comparison schools that used alternative reform models, *Accelerated Schools* (Levin, 1987) and the *School Development Program* (Comer et al., 1996). The schools were matched on demographic factors in a matched post-hoc comparison. Third graders (N=228) in seven CMCD schools were individually matched with students in seven control schools (N=228) based on their second grade Stanford-9 scores, taken in 1998. Pretest scores in math were nearly identical (ES= +0.02, n.s.). At posttest, a year later, CMCD students scored significantly and substantially higher (ES= +0.53, p <.001).

Across three post-hoc matched studies, the median effect size for CMCD was +0.40.

Supplemental Programs

In all of the instructional process programs discussed above, time for instruction was the same in experimental and control groups. However, there are several approaches that supplement core classroom supplemental instruction, taking place either during time scheduled for math or in additional time. The most common of these are computer-assisted instruction programs, discussed earlier in this paper. This section discusses four diverse approaches that have in common the fact that they add instruction beyond that ordinarily allocated to mathematics: Small-group tutoring, *Accelerated Math, Every Day Counts,* and *Project SEED.*

Small-Group Tutoring

Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett (2005) developed and evaluated a small group tutoring intervention for first graders who are struggling in mathematics. The students were taught in groups of 2-3 for 30 minutes three times a week, plus 10 minutes of work on computers to build math facts skills. Lessons made extensive use of manipulatives and frequent assessment.

The evaluation took place over 16 weeks in ten schools in a southeastern city. Across 41 first grade classrooms, 139 students were identified as at risk based on a battery of individually administered tests. These students were randomly assigned to tutoring or non-tutoring conditions. Students in these two groups were well matched on demographics, math test scores, and ability. Outcomes varied by outcome measures. Significant differences favoring the tutored students were found on Woodcock-Johnson Calculations (ES= +0.56, p<.05), but there were no differences on Woodcock-Johnson Applied Problems (ES= +0.09, n.s.).

Every Day Counts

Every Day Counts (Great Source, 2006) is an interactive K-6 bulletin board program designed to supplement ordinary math instruction with discussions about math concepts built around the calendar and other classroom routines. The evaluation involved 587 fifth graders from 28 classes in 13 high-poverty schools in New Haven, CT. Schools were initially assigned at random, but the addition of non-randomly assigned classes makes this a matched study.

In a six-month study, students were pre- and posttested on a measure patterned on the Connecticut Mastery Test (CMT). Differences were modest, but significantly favored the *Every Day Counts* classes (ES=+0.15, p<.05).

Project SEED

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Project SEED is a supplementary mathematics program in which university mathematicians and scientists teach elementary students high-level mathematics concepts. The intention of the program is both to help students develop their math skills and to motivate them to continue their education in mathematics into middle and high school. The instruction focuses on questions to students designed to get them to think creatively and productively in mathematics. *Project SEED* adds an extra daily math period to the school schedule, so effects of the program may be partly or entirely accounted for by this additional instruction time, but this is nevertheless an interesting approach to math reform.

Several evaluations of *Project SEED* took place in the Detroit Public Schools. One study followed children in grades 4-6 who had had one, two, or three semesters of *Project SEED* instruction (Webster, 1994). The 83 students who experienced one semester of *Project SEED* each year for three years were compared to individually-matched students who were in *Project SEED* schools but did not experience *Project SEED*, and 83 students in other Detroit schools, matched to the *Project SEED* students in SES and CAT pretests, among other variables. Students were pre- and posttested on the California Achievement Test (CAT). Results at the end of sixth grade indicated that *Project SEED* students scored substantially higher on CAT math than the non-*SEED* students in *SEED* schools (ES=+0.75, p<.01) or those in non-*SEED* schools (ES=+0.71, p<.01). Analyses showed smaller benefits of participation for one semester (mean ES=+0.27, p<.01), and two semesters (mean ES=+0.68, p<.01).

A later Detroit study (Webster & Dryden, 1998) compared 302 third graders who had had at least 14 weeks of *Project SEED* to 302 matched control students in gains over a year on the MAT6. Once again, analyses of covariance showed the *Project SEED* students to have made significantly greater gains (ES=+0.25, p<.01).

Conclusions: Instructional Process Strategies

Research on instructional process strategies tends to be of much higher quality than research on mathematics curricula or computer-assisted instruction. Out of 36 studies, 19 used randomized or randomized quasi-experimental designs. Many had small samples and/or short durations, and in some cases there were confounds between treatments and teachers, but even so, these are relatively high-quality studies, most of which were published in peer-reviewed journals. The median effect size for randomized studies was +0.33, and the median was also +0.33 for all studies taken together.

The research on instructional process strategies identified several methods with strong positive outcomes on student achievement. In particular, the evidence supports various forms of cooperative learning, especially *Classwide Peer Tutoring* and *PALS*, which are pair learning methods, and *Student Teams-Achievement Divisions,* and *TAI Math*, which use groups of four. Project CHILD, which also uses cooperative learning, was successfully evaluated in one study.

Two programs that focus on classroom management and motivation also had strong evidence of effectiveness in large studies. These are the *Missouri Mathematics Project*, with three large randomized and randomized quasi-experiments with positive outcomes, and *Consistency Management & Cooperative Discipline*. Positive effects were also seen for *Dynamic*

Pedagogy and *Cognitively Guided Instruction*, which focus on helping teachers understand math content and pedagogy. Four small studies supported *Direct Instruction* models, *Connecting Math Concepts (CMC),* and *User-Friendly Direct Instruction*.

Programs that supplemented traditional classroom instruction also had strong positive effects. These include small group tutoring for struggling first graders and *Project SEED*, which provides a second math period focused on high-level math concepts.

The research on these instructional process strategies suggests that the key to improving math achievement outcomes is changing the way teachers and students interact in the classroom. It is important to be clear that the well-supported programs are not ones that just provide generic professional development, or professional development focusing on mathematics content knowledge alone. What characterizes the successfully evaluated programs in this section is a focus on how teachers use instructional process strategies, such as using time effectively, keeping children productively engaged, giving children opportunities and incentives to help each other learn, and motivating students to be interested in learning mathematics.

Overall Patterns of Outcomes

Across all categories, 87 studies met the inclusion criteria for this review, of which 36 used randomized (R) or randomized quasi-experimental (RQ) designs: 13 studies (2R) of mathematics curricula, 38 (11R, 4RQ) of CAI, and 36 (9R, 10RQ) of instructional process programs. Across all studies, the median effect size was +0.22. The effect size for randomized and randomized quasi-experimental studies (N=36) was +0.29, and for fully randomized studies (N=22) it was +0.28, indicating that higher-quality studies generally produced effects similar to those of matched quasi-experimental studies. Recall that the matched studies had to meet stringent methodological standards, so the similarity between randomized and matched outcomes reinforces the observation made by Glazerman, Levy, & Myers (2002) and Torgerson (2006) that high-quality studies with well-matched control groups produce outcomes similar to those of randomized experiments.

Overall effect sizes differed, however, by type of program. Median effect sizes for all qualifying studies were +0.10 for mathematics curricula, +0.19 for CAI programs, and +0.33 for instructional process programs. Effect sizes were above the overall median (+0.22) in 15% of studies of mathematics curricula, 37% of CAI studies, and 72% of instructional process programs. The difference in effect sizes between the instructional process and other programs is statistically significant (χ^2=15.71, p<.001).

With only a few exceptions, effects were similar for disadvantaged and middle class students, and for students of different ethnic backgrounds. Effects were also generally similar on all subscales of math tests, except that CAI and *TAI Math* generally had stronger effects on measures of computations than on measures of concepts and applications.

Summarizing Evidence of Effectiveness for Current Programs

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

In several recent reviews of research on outcomes of various educational programs, reviewers have summarized program outcomes using a variety of standards. This is not as straightforward a procedure as might be imagined, as several factors must be balanced (Slavin, 2007). These include the number of studies, average effect sizes, and methodological quality of studies.

The problem is that the number of studies of any given program is likely to be small, so simply averaging effect sizes (as in meta-analyses) is likely to overemphasize small, biased, or otherwise flawed studies with large effect sizes. For example, in the current review there are several very small matched studies with effect sizes in excess of +1.00, and these outcomes cannot be allowed to overbalance large and randomized studies with more modest effects. Emphasizing numbers of studies can similarly favor programs with many small, matched studies, which may collectively be biased toward positive findings by "file drawer effects." The difference in findings for CAI programs between small numbers of randomized experiments and large numbers of matched experiments shows the danger of emphasizing numbers of studies without considering quality. Finally, emphasizing methodological factors alone risks eliminating most studies or emphasizing studies that may be randomized but are very small, confounding teacher and treatment effects, or may be brief, artificial, or otherwise not useful for judging the likely practical impact of a treatment.

In the present review, we applied a procedure for characterizing the strength of the evidence favoring each program that attempts to balance methodological, replication, and effect size factors. Following the What Works Clearinghouse (2006), CSRQ (2005), and Borman et al. (2003), but placing a greater emphasis on methodological quality, we categorized programs as follows (see Slavin, 2007):

Strong Evidence of Effectiveness

At least one large randomized or randomized quasi-experimental study, or two smaller studies, with a median effect size of at least +0.20. A large study is defined as one in which at least ten classes or schools, or 250 students, were assigned to treatments.

Moderate Evidence of Effectiveness

At least one randomized or randomized quasi-experimental study, or a total of two large or four small qualifying matched studies, with a median effect size of at least +0.20.

Limited Evidence of Effectiveness

At least one qualifying study of any design with a statistically significant effect size of at least +0.10.

Insufficient Evidence of Effectiveness

One or more qualifying study of any design with a nonsignificant outcome and a median effect size less than +0.10.

N No Qualifying Studies

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

================

Table 4 Here

================

Table 4 summarizes currently available programs falling into each of these categories (within categories, programs are listed in alphabetical order). Note that programs that are not currently available, primarily the older CAI programs, do not appear in the table, as it is intended to represent the range of options from which today's educators might choose.

In line with the previous discussions, the programs represented in each category are strikingly different. In the "Strong Evidence" category appear five instructional process programs, four of which are cooperative learning programs: *Classwide Peer Tutoring, PALS, Student Teams-Achievement Divisions,* and *TAI Math.* The fifth program is a classroom management and motivation model, the *Missouri Mathematics Program*.

The "Moderate Evidence" category is also dominated by instructional process programs, including two supplemental designs, small-group tutoring and *Project SEED*, as well as *Cognitively Guided Instruction*, which focuses on training teachers in mathematical concepts, and *Consistency Management & Cooperative Discipline*, which focuses on school and classroom management and motivation. *Connecting Math Concepts (CMC)*, an instructional process program tied to a specific curriculum, also appears in this category. The only current CAI program with this level of evidence is *Classworks*.

The "Limited Evidence" category includes five math curricula, *Everyday Mathematics, Excel Math, Growing with Mathematics, Houghton Mifflin Mathematics*, and *Knowing Mathematics. Dynamic Pedagogy, Project CHILD,* and *Mastery Learning*, instructional process programs, are also in this category, along with *Lightspan* and *Accelerated Math*. Four programs, listed under "insufficient evidence of effectiveness," had only one or two studies, which failed to find significant differences. The final category, "no qualifying studies," lists 48 programs.

Discussion

The research reviewed in this article evaluates a broad range of strategies for improving mathematics achievement. Across all topics, the most important conclusion is that there are fewer high-quality studies than one would wish for. Although a total of 87 studies across all programs qualified for inclusion, there were small numbers of studies on each particular program. There were 36 studies, 19 of which involved instructional process strategies, that randomly assigned schools, teachers, or students to treatments, but many of these tended to be small and therefore to confound treatment effects with school and teacher effects. There were several large scale, multi-year studies, especially of mathematics curricula, but these tended to be post-hoc matched quasi-experiments, which can introduce serious selection bias. Clearly, more randomized evaluations of programs used on a significant scale over a year or more are needed.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

This being said, there were several interesting patterns in the research on elementary mathematics programs. One surprising observation is the lack of evidence that it matters very much which textbook schools choose (median ES=+0.10 across 13 studies). Quality research is particularly lacking in this area, but the mostly matched post-hoc studies that do exist find modest differences between programs. NSF-funded curricula such as *Everyday Mathematics, Investigations*, and *Math Trailblazers* might have been expected to at least show significant evidence of effectiveness for outcomes such as problem-solving or concepts and applications, but the quasi-experimental studies that qualified for this review find little evidence of strong effects even in these areas. The large national study of these programs by Sconiers et al. (2003) found effect sizes of only +0.10 for all outcomes, and the median effect size for five studies of NSF-funded programs was +0.12.

It is possible that the state assessments used in the Sconiers et al. (2003) study and other studies may have failed to detect some of the more sophisticated skills taught in NSF-funded programs but not other programs, a concern expressed by Schoenfeld (2006) in his criticism of the What Works Clearinghouse. However, in light of the small effects seen on outcomes such as problem solving, probability and statistics, geometry, and algebra, it seems unlikely that misalignment between the NSF-sponsored curricula and the state tests account for the modest outcomes.

Studies of computer-assisted instruction found a median effect size (ES=+0.19) higher than that found for mathematics curricula, and there were many more high-quality studies of CAI. A number of studies showed substantial positive effects of using CAI strategies, especially for computations, across many types of programs. However, the highest-quality studies, including the few randomized experiments, mostly find no significant differences.

CAI effects in math, although modest in median effect size, are important in light of the fact that in most studies, CAI was used for only about 30 minutes three times a week or less. The conclusion that CAI is effective in math is in accord with the findings of a recent review of research on technology applications by Kulik (2003), who found positive effects of CAI in math but not reading.

The most striking conclusion from the review, however, is the evidence supporting various instructional process strategies. Twenty randomized experiments and randomized quasi-experiments found impressive impacts (median ES=+0.29) of programs that target teachers' instructional behaviors rather than math content alone. Several categories of programs were particularly supported by high-quality research. Cooperative learning methods in which students work in pairs or small teams and are rewarded based on the learning of all team members were found to be effective in nine well-designed studies, eight of which used random assignment, with a median effect size of +0.29. These included studies of *Classwide Peer Tutoring, PALS,* and *Student Teams-Achievement Divisions. Team Accelerated Instruction (TAI),* which combines cooperative learning and individualization, also had strong evidence of effectiveness. Another well-supported approach included programs that focus on improving teachers' skills in classroom management, motivation, and effective use of time, in particular the *Missouri Mathematics Project* and *Consistency Management & Cooperative Discipline.* Studies supported programs focusing on helping teachers introduce mathematics concepts effectively, such as *Cognitively Guided Instruction, Dynamic Pedagogy,* and *Connecting Math Concepts.*

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Supplementing classroom instruction with well-targeted supplementary instruction is another strategy with strong evidence of effectiveness. In particular, small-group tutoring for first graders struggling in math and *Project SEED*, which provides an additional period of instruction from professional mathematicians, have strong evidence.

The debate about mathematics reform has focused primarily on curriculum, not on professional development or instruction (see, for example, AAAS, 2000; NRC, 2004). Yet this review suggests that in terms of outcomes on traditional measures, such as standardized tests and state accountability assessments, curriculum differences appear to be less consequential than instructional differences. This is not to say that curriculum is unimportant. There is no point in teaching the wrong mathematics. The research on the NSF-supported curricula is at least comforting in showing that reform-oriented curricula are no less effective than traditional curricula on traditional measures, and may be somewhat more effective, so their contribution to non-traditional outcomes does not detract from traditional ones. The movement led by NCTM to focus math instruction more on problem solving and concepts may account for the gains over time on NAEP, which itself focuses substantially on these domains.

Also, it is important to note that the three types of approaches to mathematics instruction reviewed here do not conflict with each other, and may have additive effects if used together. For example, schools might use an NSF-supported curriculum such as *Everyday Mathematics* with well-structured cooperative learning and supplemental computer-assisted instruction, and the effects may be greater than those of any of these programs by itself. However, the findings of this review suggest that educators as well as researchers might do well to focus more on how mathematics is taught, rather than expecting that choosing one or another textbook by itself will move their students forward.

As noted earlier, the most important problem in mathematics education is the gap in performance between middle and lower class students and between White and Asian-American students and African American, Hispanic and Native American students. The studies summarized in this review took place in widely diverse settings, and several of them reported outcomes separately for various subgroups. Overall, there is no clear pattern of differential effects for students of different social class or ethnic backgrounds. Programs found to be effective with any subgroup tend to be effective with all groups. Rather than expecting to find programs with different effects on students in the same schools and classrooms, the information on effective mathematics programs might better be used to address the achievement gap by providing research-proven programs to schools serving many disadvantaged and minority students. Federal Reading First and Comprehensive School Reform programs were intended to provide special funding to help high-poverty, low-achieving schools adopt proven programs. A similar strategy in mathematics could help schools with many students struggling in math to implement innovative programs with strong evidence of effectiveness, as long as the schools agree to participate in the full professional development process used in successful studies and to implement all aspects of the program with quality and integrity.

The mathematics performance of America’s students does not justify complacency. In particular, schools serving many students at risk need more effective programs. This article provides a starting place in determining which programs have the strongest evidence bases today. Hopefully, higher quality evaluations of a broader range of programs will appear in the coming years. What is important is that we use what we know now at the same time as we work to improve our knowledge base in the future, so that our children receive the most effective mathematics instruction we can give them.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

References

AAAS (2000). *Middle school mathematics textbooks: A benchmark-based evaluation*. Washington, DC: Author.

Abram, S.L. (1984). *The effect of computer assisted instruction on first grade phonics and mathematics achievement computation*. Unpublished doctoral dissertation, Northern Arizona University.

Abt Associates, Inc. (Undated). Independent evaluation of the effectiveness of the math steps curriculum (Houghton Mifflin). Unpublished manuscript.

Adams, G. & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR.* Seattle, WA: Educational Achievement Systems.

Ai, X. (2002). *District Mathematics Plan evaluation: 2001-2002 evaluation report.* Program Evaluation and Research Branch: Los Angeles Unified School District, CA

Al-Halal, A. (2001). *The effects of individualistic learning and cooperative learning strategies on elementary students' mathematics achievement and use of social skills.* Unpublished doctoral dissertation, Ohio University.

Allinder, R. M. & Oats, R. G. (1997). Effects of Acceptability on Teachers' Implementation of Curriculum-Based Measurement and Student Achievement in Mathematics Computation. *Remedial and Special Education, 18* (2), 113-120.

Anderson, L., Scott, C., & Hutlock, N. (1976, April). *The effects of a mastery learning program on selected cognitive, affective and ecological variables in grades 1 through 6*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Armour-Thomas, E., Walker, E., Dixon-Roman, E.J., Mejia, B., & Gordon, E.J. (2006). *An evaluation of Dynamic Pedagogy on third-grade math achievement of ethnic minority children.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Atkeison-Cherry, N. K. (2004). *A comparative study of mathematics learning of third-grade children using a scripted curriculum and third-grade children using a non-scripted curriculum.* Unpublished Doctoral Dissertation, Union University.

Austin Independent School District. (2001). *Austin collaborative for mathematics education, 1999-2000 evaluation*. Unpublished manuscript.

Autry, S.L. (2001). *First-grade students' achievement and attitude using two different approaches in mathematics instruction.* Unpublished doctoral dissertation, University of Alabama.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Axelrod, S., McGregor, G., Sherman, J., Hamlet, C. (Fall 1987).Effects of Video Games as Reinforcers for Computerized Addition Performance. *Journal of Special Education Technology, 9* (1), 1-8.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research teaching mathematics to low-achieving students. *The Elementary School Journal, 103* (1), 51-69.

Ball, D.L. (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education.* Santa Monica, CA: RAND.

Bass, G., Ries, R., and Sharpe, W. (1986). Teaching basic skills though microcomputer assisted instruction. *Journal of Educational Computing Research. 2* (2), 207-219.

Baxter, J. A., Woodward, J., Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms. *The Elementary School Journal, 101*(5), 529-547.

Beck Evaluation & Testing Associates, Inc. (2006). *Progress in Mathematics 2006: Grade 1 pre-post field test evaluation*. Pleasantville, NY: Author.

Becker, H.J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8* (1), 1-41.

Becker, H.J. (1992). A model for improving the performance of integrated learning systems: Mixed individualized/group/whole class lessons, cooperative learning, and organizing time for teacher-led remediation of small groups. *Educational Technology, 32* (9), 6-15.

Becker, H.J. (1994). Mindless or mindful use of integrating learning systems. *International Journal of Educational Research, 21* (1), 65-79.

Becker, W.C. & Gersten, R. (1982). A follow-up of Follow Through: The later effects of the Direct Instruction Model on children in the fifth and sixth grades. *American Educational Research Journal, 19* (1), 75-92.

Bedell, J.P. (1998). *Effects of reading and mathematics software formats on elementary students' achievement.* Unpublished doctoral dissertation, University of Miami.

Bereiter, C. & Kurland, M. (1981-1982). A constructive look at Follow-Through results. *Interchange, 12*, 1-22.

Birch, J.H. (2002). *The effects of the Delaware Challenge Grant Program on the standardized reading and mathematics test scores of second and third grade students in the Caesar Rodney School District*. Unpublished doctoral dissertation, Wilmington College.

Block, J., & Anderson, L. (1975). *Mastery learning in classroom instruction*. New York: MacMillan.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Block, J., & Burns, R. (1976). Mastery learning. In L.S. Shulman (Ed.), *Review of Research in Education* (Vol. 4, pp. 3-49). Itasca, IL: F.E. Peacock.

Bolser, S. & Gilman, D. (2003). *Saxon Math, Southeast Fountain Elementary School: Effective or ineffective?* (ERIC Document Reproduction Service No. ED 474 537).

Borton, W.M. (1988). The effects of computer managed mastery learning on mathematics test scores in the elementary school. *Journal of Computer-Based Instruction, 15* (3), 95-98.

Bosfield, G. (2004). *A comparison of traditional mathematical learning and cooperative mathematical learning*. Unpublished Masters Thesis, California State University Dominguez Hills.

Boys, C.J. (2003). *Mastery orientation through task-focused goals: Effects on achievement and motivation.* Unpublished master's thesis, University of Minnesota.

Brandt, W.C., & Hutchinson, C. (2005). *Romulus Community Schools comprehensive school reform evaluation.* Naperville, IL: Learning Point Associates.

Brem, S.K. (2003). *AM users outperform controls when exposure and quality of interaction are high: A two-year study of the effects of Accelerated Math and Math Renaissance performance in a Title I elementary school (Technical report).* Tempe: Arizona State University.

Brent, G. & DiObilda, N. (1993). Effects of curriculum alignment versus Direct Instruction on urban children. *Journal of Educational Research, 86* (6), 333-338.

Briars, D. J. (2004, July). *The Pittsburgh story: Successes and challenges in implementing standards-based mathematics programs*. Paper presented at the meeting of the UCSMP Everyday Mathematics Leadership Institute, Lisle, IL.

Briars, D., & Resnick, L. (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Brown, F., & Boshamer, C.C. (2000). Using computer assisted instruction to teach mathematics: A study. *The Journal of the National Alliance of Black School Educators, 4* (1), 62-71.

Brush, T. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Research and Development, 45* (1), 51-64.

Brush, T.A., Armstrong, J., Barbrow, D., & Ulintz, L. (1999). Design and delivery of integrated learning systems: Their impact on student achievement and attitudes. *Journal of Educational Computing Research, 21* (4), 475-486.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Bryant, R.R. (1981). *Effects of Team-Assisted Individualization on the attitudes and achievement of third, fourth, and fifth grade students of mathematics.* Unpublished doctoral dissertation, University of Maryland.

Burke, A.J. (1980). *Students' potential for learning contrasted under tutorial and group approaches to instruction.* Unpublished doctoral dissertation, University of Chicago.

Burkhouse, B., Loftus, M., Sadowski, B., & Buzad, K. (2003). *Thinking Mathematics as professional development: Teacher perceptions and student achievement.* Scranton, Pennsylvania: Marywood University. (ERIC Document Reproduction Service No. ED 482 911).

Burton, S.K. (2005). *The effects of ability grouping on academic gains of rural elementary school students.* Unpublished doctoral dissertation, Union University.

Butzin, S. M. (2001). Using instructional technology in transformed learning environments: An evaluation of Project CHILD. *Journal of Research on Computing in Education, 33*, 367-373.

Butzin, S., King, F. J. (1992). An evaluation of Project CHILD. *Florida Technology in Education Quarterly, 4* (4), 45 - 63.

Calvery, R., Bell, D., and Wheeler, G. (1993, November). A comparison of selected second and third graders' math achievement: Saxon vs Holt. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA.

Campbell, P.F. (1996). Empowering children and teachers in the elementary mathematics classrooms of urban schools. *Urban Education, 30* (4), 449-475.

Campbell, P.F., Rowan, T.E., & Cheng, Y. (1995). *Project IMPACT: Mathematics achievement in predominantly minority elementary classrooms attempting reform.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Cardelle-Elawar, M. (1990). Effects of feedback tailored to bilingual students' mathematics needs on verbal problem solving. *The Elementary School Journal, 91* (2), 165-175.

Cardelle-Elawar, M (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching & Teacher Education*, *8* (2), 109-121.

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11* (1), 81-95.

Carpenter, T. P. Fennema, E., Franke, M. L., Empson, S. B., & Levi, L. W. (1999). *Children's mathematics: Cognitively Guided Instruction*. Portsmouth, NH: Heinemann.

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 Education Research Journal, 26,* 499-531.

Carrier, C., Post, T., & Heck, W. (1985). Using microcomputers with fourth-grade students to reinforce arithmetic skills. *Journal for Research in Mathematics Education, 16* (1), 45-51.

Carroll, W. M. (1993). *Mathematical knowledge of kindergarten and first-grade students in Everyday Mathematics.* UCSMP Report. Unpublished manuscript.

Carroll, W. M. (1994-1995). *Third grade everyday mathematics students' performance on the 1993 and 1994 Illinois state mathematics test.* Unpublished manuscript.

Carroll, W. (1995). *Report on the field test of Fifth Grade Everyday Mathematics.* Chicago, IL: University of Chicago

Carroll, W. M. (1996). *A follow-up to the fifth-grade field test of Everyday Mathematics: Geometry, and mental and written computation.* Chicago. University of Chicago School Mathematics Project.

Carroll, W. M. (1996b). Use of invented algorithms by second graders in a reform mathematics curriculum. *Journal of Mathematical Behavior, 15* (2), 137-150.

Carroll, W.M. (1996c). Mental computation of students in a reform-based mathematics curriculum. *School Science and Mathematics, 96* (6), 305-311.

Carroll, W. M. (1997). Mental and written computation: Abilities of students in a reform-based mathematics curriculum. *The Mathematics Educator, 2* (1), 18-32.

Carroll, W. (1998). Geometric knowledge of middle school students in a reform-based mathematics curriculum. *School Science and Mathematics, 98* (4), 188-197.

Carroll, W.M. (2000). Invented computational procedures of students in a standards-based curriculum. *Journal of Mathematical Behavior, 18* (2), 111-121.

Carroll, W. M. (2001). *A longitudinal study of children in the everyday mathematics curriculum.* Unpublished manuscript.

Carroll, W. M. (2001). Students in a standards-based curriculum: Performance on the 1999 Illinois state achievement test. *Illinois Mathematics Teacher, 52* (1), 3-7.

Carroll, W. M., & Fuson, K. C. (1998). Multidigit computation skills of second and third graders in everyday mathematics: A follow-up to the longitudinal study. Unpublished manuscript.

Carroll, W. M. & Isaacs, A. (2003). Achievement of students using the University of Chicago School Mathematics Project's Everyday Mathematics. In S. L Senk & D. R. Thompson

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

(Eds.) *Standards-based school mathematics curriculum: What are they? What do students learn?* (pp. 79-108). Mahwah, NJ: Laurence Erlbaum Associates.

Carrol, W.M. & Porter, D. (1994). *A field test of Fourth Grade Everyday Mathematics: Summary report.* Chicago: University of Chicago School Mathematics Project, Elementary Component.

Carter, A., Beissinger, J., Astrida, C., Gartzman, M., Kelso, C., & Wagreich, P. (2003). Student learning and achievement with math. In S. L Senk & D. R. Thompson (Eds.) *Standards-based school mathematics curriculum: What are they? What do students learn?* (pp. 45-78). Mahwah, NJ: Laurence Erlbaum Associates.

Cavanagh, S. (2006, February 15). White House suggests model used in reading to elevate math skills. *Education Week*, pp. 14, 16.

Chan, L., Cole, P., & Cahill, R. (1988). Effect of cognitive entry behavior, mastery level, and information about criterion on third graders' mastery of number concepts. *Journal for Research in Mathematics Education, 19* (5), 439-448.

Chang, K. E., Sung, Y. T., Lin, S. F. (2006). Computer-assisted learning for mathematical problem solving. *Computers & Education, 46* (2), 140-151.

Chambers, E. A. (2003). *Efficacy of educational technology in elementary and secondary classrooms: A meta-analysis of the research literature from 1992-2002.* Unpublished doctoral dissertation, Southern Illinois University at Carbondale.

Chase, A., Johnston, K., Delameter, B., Moore, D., & Golding, J. (2000). *Evaluation of The Math Steps curriculum: Final report.* Boston, MA: Houghton-Mifflin.

Chiang, A. (1978) *Demonstration of the use of computer-assisted instruction with handicapped children: Final report.* (Report No. 446-AM-66076A). Arlington, VA: RMC Research Corp. (ERIC Document Reproduction Service No. ED 116 913.)

Clariana, R.B. (1996). Differential achievement gains for mathematics computation, concepts, and applications with an integrated learning system. *The Journal of Computers in Mathematics and Science, 15* (3), 203-215.

Clariana, R.B. (1994). *The effects of an integrated learning system on third graders' mathematics and reading achievement.* San Diego, CA: Jostens Learning Corporation. (ERIC Document Reproduction Service No. ED 409 181).

Clark, R.E. (Ed.) (2001). *Learning from media: Arguments, analysis, and evidence.* Greenwich, CT: Information Age.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Clayton, I.L. (1992). *The relationship between computer-assisted instruction in reading and mathematics achievement and selected student variables.* Unpublished doctoral dissertation, University of Southern Mississippi.

Clements, D.H., & Battista, M.T. (1990). The effects of logo on children's conceptualizations of angle and polygons. *Journal for Research in Mathematics Education, 21* (5), 356-371.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22* (1), 3-29.

Cooper, H. (1998). *Synthesizing research (3rd ed.).* Thousand Oaks, CA: Sage.

Cooperative Mathematics Project (1996). *Findings from the evaluation of the Number Power Mathematics Program: Final report to the National Science Foundation*. Oakland, CA: Developmental Studies Center.

Cox, B.F. (1986). *Instructional management system: How it affects achievement, self-concept, and attitudes toward math of fifth-grade students.* Unpublished doctoral dissertation, Oklahoma State University.

Cramer, K.A., Post, T.R., & delMas, R.C. (2002). Initial fraction learning by fourth and fifth grade students: A comparison of the effects of using commercial curricula with the effects of using the Rational Number Project curriculum. *Journal for Research in Mathematics Education, 33* (2), 111-144.

Crawford, D.B. & Snider, V.E. (2000). Effective mathematics instruction: The importance of curriculum. *Education and Treatment of Children, 23*(2), 122-142.

Crenshaw, H.M. (1982). *A comparison of two computer assisted instruction management systems in remedial math.* Unpublished doctoral dissertation, Texas A&M University.

Dahn, V.L. (1992). *The effect of integrated learning systems on mathematics and reading achievement and student attitudes in selected Salt Lake City, Utah, elementary schools.* Unpublished doctoral dissertation, Brigham Young University.

Dalton, D. W. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. *Journal of Computer-Based Instruction, 17* (1), 8-16.

Dalton, D. W., Hannafin, M. J., & Hooper, S. (1989). Effects of individual and cooperative computer-assisted instruction on student performance and attitudes. *Educational Technology Research and Development, 37* (2), 15-24.

Davidson, N. & Kroll, D. L. (1991). An overview of research on cooperative learning related to mathematics. *Journal for Research in Mathematics Education, 22* (5), 362-365.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

De Russe, J.T. (1999). *The effect of staff development training and the use of cooperative learning methods on the academic achievement of third through sixth grade Mexican-American students.* Unpublished doctoral dissertation, Texas A&I University.

Dilworth, R. & Warren, L. (1980). *An independent investigation of Real Math: The field-testing and learner-verification studies.* San Diego, California: Center for the Improvement of Mathematics Education.

Ding, C. S., Davison, M. L. (2004). A longitudinal study of math achievement gains for initially low achieving students. *Contemporary Educational Psychology, 30,* 81-95.

Donnelly, L. (2004, April). *Year two results: Evaluation of the implementation and effectiveness of SuccessMaker during 2002-2003.* Paper presented at the annual meeting of the American Educational Research Association, (San Diego, California).

Dorsey, O.L. & Burleson, J. (1982). *Automated learning, individual instruction, and computers in the small school classroom.* San Marcos, TX: Southwest Texas State University.

Drueck, J. V., Fuson, K. C., Carroll, W. M., and Bell, M. S. (1995, April 20-24). *Performance of U.S. first graders in a reform math curriculum compared to Japanese, Chinese, and traditionally taught U.S. students*. Paper presented at the Annual Meeting of the American Education Research Association, San Francisco, CA.

DuPaul, G.J., Ervin, R.A., Hook, C.L., & McGoey, K.E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: Effects on classroom behavior and academic performance. *Journal of Applied Behavior Analysis, 31,* 595-592.

Earnheart, M. (1989). *A study of the effectiveness of mastery learning in a low socioecomic setting.* Unpublished doctoral dissertation, University of Mississippi.

EDSTAR, Inc. (2004). *Large-scale evaluation of student achievement in districts using Houghton Mifflin Mathematics: Phase two*. Raleigh, NC: EDSTAR.

Ellington, A. J. (2001). *Effects of hand-held calculators on precollege students in mathematics classes: A meta-analysis.* Unpublished doctoral dissertation, University of Tennessee.

Emihovich, C & Miller, G. (1988). Effects of Logo and CAI on black first graders' achievement, reflectivity, and self-esteem. *The Elementary School Journal, 88* (5), 472-487.

Engelmann, S., Carnine, D., Kelly, B., & Engelmann, O. (1991-1995). *Connecting Math Concepts.* Chicago: Science Research Associates.

Erdner, R.A., Guy, R.F., & Bush, A. (1998). The impact of a year of computer-assisted instruction on the development of first grade learning skills. *Journal of Educational Computing Research, 18*(4), 369-386.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Estep, S.G. (1997). An investigation of the relationship between integrated learning systems and academic achievement. *Journal of Educational Technology Systems, 28* (1), 5-19.

Estep, S. G., McInerney, W. D., Vockell, E., & Kosmoski, G. (1999-2000). An investigation of the relationship between integrated learning systems and academic achievement. *Journal of Educational Technology Systems*, *28*(1), 5-19.

Fahsl, A.J. (2001). *An investigation of the effects of exposure to Saxon math textbooks, socioeconomic status and gender on math achievement scores.* Unpublished doctoral dissertation, Oklahoma State University.

Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement. *Journal of Educational Psychology, 87* (2), 272-281.

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.

Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of Reciprocal Peer Tutoring across elementary school settings. *Journal of School Psychology*, *28*, 309-323.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27* (4), 403-434.

Flaherty, J., Connolly, B., & Lee-Bayha, J. (2005). *Evaluation of the First In Math® Online Mathematics Program.* San Diego, WestEd.

Fletcher, J.D., Hawley, D.E., & Piele, P.K. (1990). Costs, effects, and utility of microcomputer-assisted instruction. *American Educational Research Journal, 27* (4), 783-806.

Fletcher-Flinn, C. M. & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research*, *12* (3), 219-241.

Florida TaxWatch's Comparative Evaluation of Project CHILD: Phase IV, 2005 http:www.floridataxwatch.org/projchild/projchild4.html

Flowers, J. (1998). *A study of proportional reasoning as it relates to the development of multiplication concepts.* Unpublished doctoral dissertation, University of Michigan.

Follmer, R. (2001). *Reading, mathematics, and problem solving: The effects of direct instruction in the development of fourth grade students' strategic reading and problem-solving approaches to text-based, non-routine mathematics problems.* Unpublished doctoral dissertation, Widener University.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in Everday Mathematics classrooms. *Journal for Research in Mathematics Education, 30* (2), 148-170.

Freiberg, H. J. (1996). From tourists to citizens in the classroom. *Educational Leadership, 54* (1), 32-36.

Freiberg, H. J. (Ed.). (1999). *Beyond Behaviorism: Changing the Classroom Management Paradigm*. Boston, MA: Allyn & Bacon.

Freiberg, H.J., Connell, M.L., & Lorentz, J. (2001). Effects of consistency management on student mathematics achievement in seven Chapter 1 elementary schools. *Journal of Education for Students Placed at Risk, 6* (3), 249-270.

Freiberg, H.J., Prokosch, N., Tresister, E.S., & Stein, T. (1990). Turning around five at-risk elementary schools. *Journal of School Effectiveness and School Improvement, 1* (1), 5-25.

Freiberg, H.J., Stein, T.A., and Huang, S. (1995). Effects of classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation, 1* (1), 36-66.

Fuchs, L., Compton, D., Fuchs, D., Paulsen, K., Bryant, J., & Hamlett, C. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97* (3), 493-513.

Fuchs, L., Fuchs, D., Finelli, R., Courey, S., & Hamlett, C. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41* (2), 419-445.

Fuchs, L., Fuchs, D., Finelli, R., Courey, S., Hamlett, C., Sones, E., & Hope, S. (2006). Teaching third graders about real-life mathematical problem solving: A randomized control study. *The Elementary School Journal, 106* (4), 293-311.

Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55* (5), 429-438.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Appleton, A.C. (2002). Explicitly teaching for transfer: Effects on the mathematical problem-solving performance of students with mathematical disabilities. *Learning Disabilities Research & Practice, 17* (2), 90-106.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35* (2), 227-267.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Hope, S. K., Hollenbeck, K. N., Capizzi, A.M., Craddock, C.F., & Brothers, B.L. (2006). Extending responsiveness-to-intervention to math problem-solving at third grade. *Teaching Exceptional Children*, 59-63.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., & Bentz (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60* (6), 518-537.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., & Karns, K. (l995). General educators' specialized adaptation for students with learning disabilities. *Exceptional Children, 61*, 440-459.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal, 97* (3), 223-249.

Fuchs, L., Fuchs, D., Hamlett, C., & Stecker, P. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28* (3), 617-641.

Fuchs, L., Fuchs, D. & Karns, K. (2001) Enhancing kindergarteners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal, 101* (5), 495-510.

Fuchs, L., Fuchs, D., Karns, K., Hamlett, C., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36* (3), 609-646.

Fuchs, L., Fuchs, D., Karns, K., Hamlett, C., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. *American Educational Research Journal, 34* (3), 513-543.

Fuchs, L.S., Fuchs, D., Karns, K., Yazdian, L., & Powell, S. (2001). Creating a strong foundation for mathematics learning with kindergarten peer-assisted learning strategies. *The Council for Exceptional Children*, *33*, 84-87.

Fuchs, L., Fuchs, D. Phillips, N., Hamlett, C., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review*, *24* (4), 604-620.

Fuchs, L., Fuchs, D., & Prentice (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37* (4), 293-306.

Fuchs, L., Fuchs, D., Prentice, K., Burch, M., Hamlett, C., Owen, R, Hosp, M., & Jancek, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95* (2), 293-305.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Fuchs, L., Fuchs, D., Prentice, K., Burch, M., Hamlett, C., Owen, R, & Schroeter, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95* (2), 306-315.

Fuchs, L., Fuchs, D., Prentice, K., Hamlett, C., Finelli, R., & Courey, S. (2004). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology, 96* (4), 635-647.

Fuchs, L., Fuchs, D., Yazdian, L., & Powell, S. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review, 31* (4), 569-583.

Fueyo, V. & Brushell, D. (1998). Using number line procedures and peer tutoring to improve the mathematics computation of low-performing first graders. *Journal of Applied Behavior Analysis, 31* (3), 417-30.

Fuson, K. C., and Carroll, W. M. (undated). *Performance of U.S. fifth graders in a reform math curriculum compared to Japanese, Chinese, and traditionally taught U.S. students.* Unpublished manuscript.

Fuson, K. C., and Carroll, W. M. (Undated). *Summary of comparison of Everyday Math (EM) and McMillan (MC): Evanston student performance on whole-class tests in grades 1, 2, 3, and 4.* Unpublished manuscript.

Fuson, K., Carroll, W., and Drueck, J. (2000). Achievement results for second and third graders using the standards-based curriculum Everyday Mathematics. *Journal for Research in Mathematics Education, 31* (3), 277-295.

Gabbert, B., Johnson, D. W., & Johnson, R. T. (2001). Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *The Journal of Psychology, 120* (3), 265-278.

Gallagher, P.A. (1991). *Training the parents of Chapter 1 math students in the use of mastery learning strategies to help increase math achievement.* Unpublished doctoral dissertation, University of Pittsburgh.

Gatti, G. (2004). ARC study. In Pearson Education, Inc. (Ed.). *Investigations in Number, Data, and Space: Validation and research*. Upper Saddle River, NJ: Pearson.

Gatti, G. (2004). *Scott Foresman-Addison Wesley Math national effect size study.* (Available from Pearson Education, K-12 School Group, 1 Lake Street, Upper Saddle River, NJ 07458)

Gentile, J. R., & Lalley, J. P. (2003). *Standards and Mastery Learning: Aligning Teaching and Assessment So All Children Can Learn.* Thousand Oaks, Calif.: Corwin.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Giancola, S. (2000). *Evaluation Results of the Delaware Challenge Grant Project Lead Education Agency: Capital School District.* Newark, DE: University of Delaware.

Gilbert-Macmillan, K. (1983). *Mathematical problem solving in cooperative small groups and whole class instruction.* Unpublished doctoral dissertation, Stanford University.

Gill, B. (1995). *Project CHILD middle school follow-up evaluation: Final report*. Daniel Memorial Institute, Jacksonville, FL.

Glassman, P. (1989). *A study of cooperative learning in mathematics, writing, and reading in the intermediate grades: A focus upon achievement , attitudes, and self-esteem by gender, race, and ability group.* Unpublished doctoral dissertation, Hofstra University.

Glazerman, S., Levy, D.M., & Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review*. Washington, DC: Mathematica.

Good, T.L. & Grouws, D.A. (1979). The Missouri Mathematics Effectiveness Project: An experimental study in fourth-grade classrooms. *Journal of Educational Psychology, 71* (3), 355-362.

Good, T.L., Grouws, D.A., & Ebmeier, H. (1983). Active mathematics teaching. New York: Longman.

Goodrow, A. (1998). *Children's construction of number sense in traditional, constructivist, and mixed classrooms*. Unpublished doctoral dissertation, Tufts University, Medford, MA.

Grandgenett, N., Jackson, R., & Willits, C. (2004). Evaluating a new mathematics curriculum: A district's multi-stakeholder approach. *NCSM Journal*, 13-21.

Graybeal, S. & Stodolsky, S. (1985). Peer work groups in elementary schools. *American Journal of Education, 93* (3), 409-428.

Great Source (2006). Every Day Counts grades K-6: Research base and program effectiveness. (Computer manual). Retrieved from http://www.greatsource.com/GreatSource/pdf/EveryDayCountsResearch206.pdf

Greenwood, C.R., Delquadri, J.C., & Hall, R.V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology, 81* (3), 371-383.

Greenwood, C.R. & Terry, B. (1993). Achievement, placement, and services: Middle school benefits of classwide peer tutoring used at the elementary school. *School Psychology Review, 22* (3), 497-517.

Griffin, S.A., Case, R., & Siegler, R.S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice*, pp. 25-49. Cambridge, MA: The MIT Press.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Grossen, B, & Ewing, S. (1994). Raising Mathematics Problem-Solving Performance: Do the NCTM Teaching Standards Help? Final Report, *Effective School Practices, 13*(2), 79-91.

Guskey, T., & Gates, S. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership, 43* (8), 73-80.

Gwaltney, L. (2000). Year three final report the Lightspan Partnership, Inc. Achieve Now Project: Unified School District 259, Wichita Public Schools. Wichita, KS:Allied Educational Research and Development Services.

Hallmark, B.W. (1994). *The effects of group composition on elementary students' achievement , self-concept, and interpersonal perceptions in cooperative learning groups.* Unpublished doctoral dissertation, University of Connecticut.

Hansen, E., & Greene, K. (2002) A recipe for math. What's cooking in the classroom: Saxon or traditional? Available: http://www.secondaryenglish.com/recipeformath.html [3/01/06].

Hativa, N. (1998). Computer-based drill and practice in arithmetic: Widening the gap between high- and low-achieving students. *American Educational Research Journal, 25* (3), 366-397.

Haynie, T.R. (1989). *The effects of computer-assisted instruction on the mathematics achievement of selected groups of elementary school students.* Unpublished doctoral dissertation, George Washington University.

Heller, L.R. & Fantuzzo, J.W. (1993). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review, 22* (3), 517-535.

Helsel, F., Hitchcock, J., Miller, G., Malinow, A., & Murray, E. (2006). *Identifying evidence-based, promising and emerging practices that use screen-based and calculator technology to teach mathematics in grades K-12: A research synthesis.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hess, R.D. & McGarvey, L.J. (1987). School-relevant effects of educational use of microcomputers in kindergarten classrooms and homes. *Journal of Educational Computing Research. 3* (3), 269-287.

Hickey, D.T., Moore, A.L., & Pellegrino, J.W. (2001). The motivational and academic consequences of elementary mathematics environments: Do contructivist innovations and reforms make a difference? *American Educational Research Journal, 38*(3), 611-652.

Hickey, D. T., Pellegrino, J. W., Goldman, S. R., Vye, N. J., & Moore, A. L. (1993, April). *Interests, attitudes, & anchored instruction: The impact of one interactive learning environment.* Paper presented at the meeting of the American Educational Research Association, Atlanta.

Hiebert, J. & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30* (2), 393-425.

Hill, H. (2004). Professional development standards and practices in elementary school math. *Elementary School Journal, 104* (3), 215-232.

Holmes, C.T., & Brown, C.L. (2003). A controlled evaluation of a total school improvement process, School Renaissance. Paper presented at the National Renaissance Conference, Nashville, TN.

Hooper, S. (1992). Effects of peer interaction during computer-based mathematics instruction. *Journal of Education Research, 85* (3), 180-189.

Houghton Mifflin (n.d.). *Knowing Mathematics Field Test Results for Lincoln Public Schools: Fall 2002-Spring 2003.* Boston, MA: Houghton Mifflin Company.

Hunter, C.T.L. (1994). *A study of the effect of instructional method on the reading and mathematics achievement of Chapter 1 students in rural Georgia*. Unpublished doctoral dissertation, South Carolina State University.

Interactive, Inc. & Metis Associates, Inc. (2002). *A multi-year analysis of the outcomes of Lightspan Achievement Now in the Cleveland Municipal School District*. Huntington, NY: Lightspan.

Interactive, Inc. (2000). Documenting the effects of Lightspan Achieve Now! in the Adams County School District 50: Year 3 report. Huntington, NY: Lightspan.

Interactive, Inc. (2001). Documenting the effects of Lightspan Achieve Now! in the Hempstead Union Free School District: Year 2 report. Huntington, NY: Lightspan.

Jamison, M. B. (2000). *The effects of linking classroom instruction in mathematics to a prescriptive delivery of lessons on an integrated learning system.* Unpublished doctoral dissertation, Texas Tech University.

Jitendra, A. K., & Hoff, K. (1996). The effects of schema-based instruction on the mathematical word-problem-solving performance of students with learning disabilities. *Journal of Learning Disabilities, 29* (4), 422-431.

Johnson, L. C. (December 1985). *The effects of the groups of four cooperative learning model on student problem-solving achievement in mathematics.* Unpublished doctoral dissertation, University of Houston.

Johnson, D.W., & Johnson, R.T. (1998). *Learning together and alone*: *Cooperative, competitive, and individualistic learning* (5th ed.). Boston: Allyn & Bacon.

Johnson, J., Yanyo, L., & Hall, M. (2002). *Evaluation of student math performance in California school districts using Houghton Mifflin mathematics*. Raleigh, NC: EDSTAR.

Johnson-Scott, P.L. (2006). *The impact of Accelerated Math on student achievement.* Unpublished doctoral dissertation, Mississippi State University.

Jung, J. (1997). *The differential effects of mathematics instruction on the achievement of students of varying performance levels*. Unpublished doctoral dissertation, University of Wisconsin, Madison.

Karper, J., & Melnick, S.A. (1993). The effectiveness of Team Accelerated Instruction on high achievers in mathematics. *Journal of Instructional Psychology, 20* (1), 49-54.

Kersh, M. (1970). *A strategy for mastery learning in fifth grade arithmetic*. Unpublished doctoral dissertation, University of Chicago.

Kelso, C.R., Booton, B., & Majumdar, D. (2003). *Washington State Math Trailblazers student achievement report.* Chicago, IL: University of Illinois at Chicago.

Kim, J. J. & Crasco, L. M. (2006). Best policies and practices in urban education reform: A summary of empirical analysis focusing on student achievements and equity. *Journal of Education for Students Placed at Risk, 11* (1), 19-37.

King, A. (1989). Verbal interaction and problem-solving within computer-assisted cooperative learning groups. *Educational Computing Research, 5* (1), 1-15.

King, F.J. & Butzin, S. (1992). An evaluation of Project CHILD. *Florida Technology in Education Quarterly, 4* (4), 45-63.

Kirk, V.C. (2003). *Investigation of the impact of integrated learning system use on mathematics achievement of elementary students.* Unpublished doctoral dissertation, East Tennessee State University.

Klein, D. (2000). *High achievement in mathematics: Lessons from three Los Angeles elementary schools.* Washington, DC: Brookings Institution Press.

Klein, D. (2005). *The state of state math standards 2005*. Washington, DC: Thomas Fordham Foundation.

Kopecky, C.L. (2005). *A case study of the Math Matters professional development program in one elementary school*. Unpublished doctoral dissertation, University of Southern California.

Kozma, R.B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42* (2), 7-19.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Kroesbergen, E.H. & Van Luit, J.E.H. (2003). Mathematical interventions for children with special education needs. *Remedial and Special Education, 24,* 97-114.

Kromhout, O.M. (1993). Evaluation report Project CHILD, 1992-1993. Tallahassee, Fl.

Kromhout, O. M. & Butzin, S. M. (1993). Integrating computers into the elementary school curriculum: An evaluation of nine Project CHILD model schools. *Journal of Research of Computing in Education, 26* (1), 55-70.

Kulik, J. A. (1994). Meta-analytic studies of findings on computerized instruction. In E. Baker & H. O'Neil (Eds.), "Technology assessment in education and training." Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Kulik, C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 71*, 75-94.

Kulik,, C., Kulik, J., & Bangert-Drowns, R. (1986, April). *Effects of testing for mastery on student learning.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Kulik,, C., Kulik, J., & Bangert-Drowns, R. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60* (2), 265-299.

Kulik, J., Kulik, C., & Bangert-Downs, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior, I,* 48-74.

Laub, C. (1995). *Computer-integrated learning system and elementary student achievement in mathematics: An evaluation study*. Unpublished Doctoral Dissertation, Temple University.

Laub, C & Wildasin, R. (1998). *Student achievement in mathematics and the use of computer-based instruction in the Hempfield School District*. Landsville, PA: Hempfield School District.

Leffler, K. (2001). Fifth-grade students in North Carolina show remarkable math gains. *Summary of Independent Math Research*. Madison, WI: Renaissance Learning.

Leiker, V. (1993). *The relationship between an integrated learning system, reading and mathematics achievement, higher order thinking skills and certain demographic variables: A study conducted in two school districts*. Unpublished doctoral dissertation, Baylor University.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Levy, M.H. (1985). *An evaluation of computer assisted instruction upon the achievement of fifth grade students as measured by standardized tests*. Unpublished Doctoral Dissertation, University of Bridgeport.

Linchevski, L. & Kutscher, B. (1998). Tell me with whom you're learning, and I'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education, 29* (5), 533-554.

Lipsey, M. W., & Wilson, D. B. (2001). Practical Meta-analysis. Thousand Oaks, CA: Sage.

Long, V.M. (1991). Effects of mastery learning on mathematics achievement and attitudes. Unpublished doctoral dissertation, University of Missouri – Columbia.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449-521.

Lykens, J.S. (2003). *An evaluation of a standards-based mathematics program (Math Trailblazers) in Delaware's Caesar Rodney School District.* Unpublished doctoral dissertation, Wilmington College.

Mac Iver, M., Kemper, E. & Stringfield, S. (2003) *The Baltimore Curriculum Project: Final report of the four-year evaluation study.* Baltimore, MD:Johns Hopkins University, Center for Social Organization of Schools.

Madden, N.A., & Slavin, R.E. (1983). Cooperative learning and social acceptance of mainstreamed academically handicapped students. *Journal of Special Education, 17*(2), 171-182.

Madden, N.A., Slavin, R.E., & Simons, K. (1999). *MathWings: Early Indicators of Effectiveness.* Baltimore, MD: Johns Hopkins University Center for Research on the Education of Students Placed at Risk.

Madden, N.A., Slavin, R.E., & Simons, K. (1999). *MathWings: Effects on student mathematics performance*. Baltimore, MD: Johns Hopkins University Center for Research on the Education of Students Placed at Risk.

Mahoney, S.J. (1990). *A study of the effects of EXCEL Math on mathematics achievement of second and fourth-grade students.* Unpublished doctoral dissertation, Northern Arizona University.

Mann, D., Shakeshaft, C., Becker, J., & Kottkamp, R. (1999). *West Virginia story: Achievement gains from a statewide comprehensive instructional technology program.* Beverly Hills, CA: Milken Family Foundation with the West Virginia Department of Education, Charleston.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Manuel, S.Q. (1987). *The relationship between supplemental computer assisted mathematics instruction and student achievement.* Unpublished doctoral dissertation, University of Nebraska.

Mathematics Evaluation Committee (1997). *Chicago mathematics evaluation report: Year two.* Pennington, New Jersey: Hopewell Valley Regional School District.

Mason, D.A. & Good, T.L. (1993). Effects of two-group and whole-class teaching on regrouped elementary students' mathematics achievement. *American Educational Research Journal, 30* (2), 328-360.

McCabe, K.J. (2001). Mathematics in our schools: An effort to improve mathematics literacy. Unpublished Masters Thesis, California State University, Long Beach.

McCormick, K. (2005). *Examining the relationship between a standards-based elementary mathematics curriculum and issues of equity.* Unpublished doctoral dissertation, Indiana University.

McDermott, P.A. & Watkins, M.W. (1983). Computerized vs. conventional remedial instruction for learning-disabled pupils. *The Journal of Special Education, 17*(1), 81-88.

McWhirt, R., Mentavlos, M., Rose-Baele, J.S., & Donnelly, L. (2003). *Evaluation of the implementation and effectiveness of SuccessMaker.* Charleston, SC: Charleston County School District.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement, 23*(3), 185-196.

Merrell, M.L.B. (1996). *The achievement and self-concept of elementary grade students who have received Direct Instruction.* Unpublished doctoral dissertation, University of Southern Mississippi.

Mevarech, Z. R. (1993). Who benefits from cooperative computer-assisted instruction? *Journal of Computing Research. 9* (4), 451-464.

Mevarech, Z.R. (1985). The effects of cooperative mastery learning strategies on mathematics achievement. *Journal of Educational Research, 78* (3), 372-377.

Mevarech, Z.R. (1991). Learning mathematics in different mastery environments. *Journal of Education Research, 84* (4), 225-231.

Mevarech, Z.R. & Rich, Y. (1985). Effects of computer-assisted mathematics instruction on disadvantaged pupils' cognitive and affective development. *Journal of Educational Research, 79*(1), 5-11.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Meyer, L. (1984). Long-term academic effects of Direct Instruction Project Follow Through. *The Elementary School Journal, 84* (4), 380-394.

Miller, B. (1991). *The effects of ability grouping and classroom climate on student achievement in selected Texas schools.* Unpublished doctoral dissertation, Texas A&M University.

Miller, H. L. (1994). Quantitative Analyses of Student Outcome Measures. *International Journal of Educational Research* 27(2), 119-36.

Mills, S.C. (1997). *Implementing integrated learning systems in elementary classrooms.* Unpublished doctoral dissertation, University of Oklahoma.

Mintz, K.S. (2000). *A comparison of computerized and traditional instruction in elementary mathematics.* Unpublished doctoral dissertation, University of Alabama.

Mokros, J., Berle-Carman, M., Rubin, A., and O'Neil, K. (1996, April 8-12). *Learning operations: Invented strategies that work.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Mokros, J. (2002). Learning to reason numerically: The impact of investigations. In S. Senk & D. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? Do students Learn?* (pp. 109-131). Mahwah, NJ: Lawrence Erlbaum Assocaites.

Mokros, J., Berle-Carman, M., Rubin, A., and Wright, T. (1994). *Full year pilot grades 3 and 4: Investigations in numbers, data, and space*. Cambridge, MA: TERC.

Monger, C.T. (1989). *Effects of a mastery learning strategy on elementary and middle school mathematics students' achievement and subject related affect.* Unpublished doctoral dissertation, University of Tulsa.

Moody, E.C. (1994). *Implementation and integration of a computer-based integrated learning system in an elementary school.* Unpublished doctoral dissertation, Florida State University.

Morgan, J.C. (1994). *Individual accountability in cooperative learning groups: Its impact on achievement and attitude with grade three students.* Unpublished master's thesis, University of Manitoba.

Moskowitz, J.M., Malvin, J.H., Schaeffer, G.A., & Schaps, E. (1983). Evaluation of a cooperative learning strategy. *American Educational Research Journal, 20* (4), 687-696.

Moss, J. & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education, 30* (2), 122-147.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). *E-DESK: A review of recent evidence on discrete educational software* (SRI International Report). Menlo Park, CA: SRI International.

National Assessment of Educational Progress (2005). *The nation's report card.* Washington, DC: National Center for Education Statistics.

National Research Council (2004). *On evaluating curricular effectiveness*. Washington, DC: National Academies Press.

Nattiv, A. (1994). Helping behaviors and math achievement gain of students using cooperative learning. *The Elementary School Journal, 94* (3), 285-297.

Nguyen, K. (1994). *The 1993-1994 Saxon mathematics evaluation report*. Oklahoma City, Oklahoma: Oklahoma City Public Schools.

Nguyen, K. & Elam, P. (1993). *The 1992-1993 Saxon mathematics evaluation report*. Oklahoma City, Oklahoma: Oklahoma City Public Schools.

Niemiec, R. P., & Walberg, H. J. (1992). The effects of computers on learning. *International Journal of Educational Research, 17*, 99-108.

Northwest Evaluation Association. (1999). *Achievement level tests technical description.* Portland, OR: Author.

Opuni, K.A. (2006). *The effectiveness of the Consistency Management & Cooperative Discipline (CMCD) model as a student empowerment and achievement enhancer: The experiences of two K-12 inner-city school systems.* Paper presented at the 4th Annual Hawaii International Conference of Education, Honolulu, Hawaii.

Orabuchi, I. I. (1992). *Effects of using interactive CAI on primary grade students' higher-order thinking skills: Inferences, generalizations, and math problem solving.* Unpublished doctoral dissertation, Texas Woman's Univeristy.

Orr, C. (1991). *Evaluating restructured elementary classes: Project CHILD summative evaluation.* Paper presented at the Southeast Evaluation Association, Tallahassee, FL.

Patterson, D. (2005). *The effects of Classworks in the classroom*. Stephenville, TX: Tarleton State University.

Pearson Education, Inc. (2004). *Investigations in Number, Data, and Space: Validation and research.* Upper Saddle River, NJ: Author.

Pellegrino, J.W., Hickey, D., Heath, A., Rewey, K., & Vye, N. J. (1992*). Assessing the outcomes of an innovative instructional program: The 1990-1991 implementation of the*

"Adventures of Jasper Woodbury." Nashville, TN: Vanderbilt University, Learning Technology Center.

Perkins, S.A. (1987). *The relationship between computer-assisted instruction in reading and mathematics achievement and selected student variables.* Unpublished doctoral dissertation, University of Michigan.

Pigott, H.E., Fantuzzo, J.W., & Clement, P.W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *Journal of Applied Behavior Analysis, 19*, 93-98.

Pratton, J., & Hales, L.W. (1986). The effects of active participation on student learning. *The Journal of Educational Research, 79* (4), 210-215.

Quinn, B., Foshay, R., & Morris, B. (2001). *Teaching early mathematics with Plato Software*. Bloomington, MN: PLATO Learning, Inc.

Quinn, D.W., & Quinn, N.W. (2001). *Evaluation series: Grades 1-8, Apache Junction Unified School District 43.* Bloomington, MN: PLATO Learning, Inc.

Quinn, D.W., & Quinn, N.W. (2001). *Evaluation series: Preliminary study: PLATO elementary math software, Fairview Elementary, Dayton, Ohio.* Bloomington, MN: PLATO Learning, Inc.

Rader, J.B. (1996). *The effect of curriculum alignment between mathematics instruction and CAI mathematics on student performance and attitudes, staff attitudes, and alignment.* Unpublished doctoral dissertation, United States International University.

Renaissance Learning, Inc. (1998a). *Accelerated Math™* (Computer Program). Wisconsin Rapids, WI: Renaissance Learning.

Renaissance Learning, Inc. (1998b). *STAR Math*. Madison, WI: Author.

Resendez, M., & Azin, M. (2005). *The relationship between using Saxon Elementary and Middle School Math and student performance on Georgia statewide assessments*. Jackson, WY: PRES Associates, Inc.

Resendez, M. & Manley, M.A. (2005). *A study on the effectiveness of the 2004 Scott Foresman-Addison Wesley elementary math program.* Jackson, WY: PRES Associates, Inc.

Resendez, M. & Sridharan, S. (2005). *A study on the effectiveness of the 2004 Scott Foresman-Addison Wesley elementary math program: Technical Report*. Jackson, WY: PRES Associates, Inc.

Resendez, M., Sridharan, S. & Azin, M. (2006). *The relationship between using Saxon elementary school math and student performance on the Texas statewide assessments*. Jackson, WY: PRES Associates, Inc.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Riordan, J., & Noyce, P. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32* (4), 368-398.

Rohrbeck, C.A., Ginsburg-Block, M.D., Fantuzzo, J.W., & Miller, T.R. (2003). Peer-assisted learning interventions wih elementary school students: A meta-analytic review. *Journal of Educational Psychology, 94* (20, 240-257.

Ross, L.G. (2003). *The effects of a standards-based mathematics curriculum on fourth and fifth grade achievement in two midwest cities.* Unpublished doctoral dissertation, University of Iowa.

Ross, S.M., & Nunnery, J.A. (2005). *The effect of School Renaissance on student achievement in two Mississippi school districts*. Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Roy, J.W. (1993). *An investigation of the efficacy of computer-assisted mathematics, reading, and language arts instruction*. Unpublished doctoral dissertation, Baylor University, Waco, TX.

Ryan, A. W. (1991). Meta-analysis of achievement effects of microcomputer applications in elementary schools. *Educational Administration Quarterly, 27*, 161-184.

Salvo, L.C. (2005). *Effects of an experimental curriculum on third graders' knowledge of multiplication facts.* Unpublished doctoral dissertation, George Mason University.

Saxon Publishers. (2001). *Mathematics results.* Norman, OK: Author.

Saxon Publishers. (2002). *Research support: Saxon Math*. Norman, OK: Author.

Schoenfeld, A.H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31* (1), 13-25.

Schoenfeld, A.H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35* (2), 13-21.

Schoenfeld, A.H. (2006). *Notes on the educational steeplechase: Hurdles and jumps in the development of research-based math instruction*. In Mark Constas and Robert Sternberg (Eds.) *Translating Educational Theory and Research into Practice*. Mahwah, NJ: Erlbaum.

Schreiber, F., Lomas, C., & Mys, D. (1988). *Evaluation of student reading and mathematics WICAT computer managed instructional program, Salina Elementary School.* Dearborn, MI: Office of Research and Evaluation, Public Schools.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Sconiers, S., Isaacs, A., Higgins, T., McBride, J., & Kelso, C. (2003). *The Arc Center tri-state student achievement study.* Lexington, MA: COMAP.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309-316.

Sharpley, A.M., Irvine, J.W., & Sharpley, C.F. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal, 20* (1), 103-111.

Shawkey, J. (1989). *A comparison of two first grade mathematics programs: Math Their Way and Explorations.* Unpublished doctoral dissertation, University of the pacific.

Sheffield, S. (2004). Evaluation of Houghton Mifflin Math © 2005 in the Chicopee Public School District: Second year results.

Sheffield, S. (2005). Evaluation of Houghton Mifflin Math © 2005 in the Chicopee Public School District: First year results.

Shonoski, L. A. (1986). *An analysis of the effectiveness of using microcomputer assisted mathematics instruction with primary and intermediate level students*. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Shyu, H. (2000). Using video-based anchored instruction to enhance learning: Taiwan's experience. *British Journal of Educational Technology 31* (1), 57–69.

Simpson, N. (2001). *Scott Foresman California Mathematics validation study pretest-posttest results* (Report No. VM-17-3005-CA). (Available from Pearson Scott Foresman, 1415 L Street, Suite 800, Sacramento, CA 95814)

Sinkis, D.M. (1993). A comparison of chapter one student achievement with and without computer-assisted instruction. Unpublished doctoral dissertation, University of Massachusetts.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15* (9), 5-11.

Slavin, R.E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*, 293-336.

Slavin, R.E. (1987). Mastery learning reconsidered. *Review of Educational Research, 57* (2), 175-213.

Slavin, R.E. (1995). *Cooperative learning: Theory, research, and practice* (2nd Ed.). Boston: Allyn & Bacon.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Slavin, R.E. (2007). *What works? Issues in synthesizing educational program evaluations.* Baltimore, MD: Center for Data-Driven Reform in Education, Johns Hopkins University.

Slavin, R.E., Hurley, E.A., & Chamberlain, A.M. (2001). Cooperative learning in schools. In N.J. Smelser and Paul B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 2756-2761). Oxford, England: Pergamon.

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., Leavey, M., & Madden, N.A. (1984). Combining cooperative learning and individuzlied instruction: Effects on student mathematics achievement, attitudes, and behaviors. *The Elementary School Journal, 84* (4), 409-422.

Slavin, R.E., & Madden, N.A. (Eds.) (2001). *One million children: Success for All*. Thousand Oaks, CA: Corwin

Slavin, R.E, Madden, N.A., & Leavey, M. (1984). Effects of cooperative learning and individualized instruction on mainstreamed students. *Exceptional Children, 50* (5), 434-443.

Slavin, R.E., & Karweit, N.L. (1985). Effects of whole class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22* (3), 351-367.

Slavin, R.E., Madden, N.A., & Leavey, M. (1984). Effects of Team Assisted Individualization on the Mathematics Achievement of academically handicapped and non-handicapped students. *Journal of Educational Psychology, 76* (5), 813-819.

Sloan, H.A. (1993). *Direct Instruction in fourth and fifth grade classrooms.* Unpublished master's thesis, Purdue University.

Snider, V.E., & Crawford, D.B. (1996). Action research: Implementing Connecting Math Concepts. *Effective School practices, 15* (2), 17-26.

Snow, M. (1993). *The effects of computer-assisted instruction and focused tutorial services on the academic achievement of marginal learners.* Unpublished doctoral dissertation, University of Miami.

Spencer, T.M. (1999). *The systemic impact of integrated learning systems on mathematics achievement*. Unpublished doctoral dissertation, Eastern Michigan University.

Spicuzza, R., & Ysseldyke, J. (1999). *Using Accelerated Math™ to enhance instruction in a mandated summer school program.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39* (6), 541-542.

SRA/McGraw-Hill (2003). *Everyday Mathematics student achievement studies: Volume 4.* Chicago: Author.

Stalling, J. (1985). A study of implementation of Madeline Hunter's model and its effects on students. *Journal of Educational Research, 78* (6), 325-337.

Stallings, J. & Krasavage, E.M. (1986). Program implementation and student achievement in a four-year Madeline Hunter Follow-Through Project. The *Elementary School Journal, 87*(2), 117-138.

Stecker, P., & Fuchs, L. (2000). Expecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15* (3), 128-134.

Steele, K.J., Battista, M.T., & Krockover, G.H. (1984). Using microcomputer-assisted mathematics instruction to develop computer literacy. *School Science and Mathematics, 84* (2), 119-125.

Stevens, J.W. (1991). *Impact of an integrated learning system on third-, fourth-, and fifth-grade mathematics achievement*. Unpublished doctoral dissertation, Baylor University, Waco, TX.

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., & Slavin, R.E. (1995). Effects of a cooperative learning approach in reading and writing on handicapped and nonhandicapped students' achievement, attitudes, and metacognition in reading and writing. *Elementary School Journal*, *95,* 241-262.

Stone, T. (1996). *The academic impact of classroom computer usage upon middle-class primary grade level elementary school children*. Unpublished doctoral dissertation, Widener University.

Sullivan, F.L. (1989). *The effects of an integrated learning system on selected fourth- and fifth-grade students.* Unpublished doctoral dissertation, Baylor University.

Suppes, P., Fletcher, J. D., Zanotti, M., Lorton, P. V., Jr., & Searle, B. W. (1973). *Evaluation of computer-assisted instruction in elementary mathematics for hearing- impaired students.* Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.

Susman, E. B. (1998). Cooperative learning: A review of factors that increase the effectiveness of cooperative computer-based instruction. *Journal of Educational Computing Research, 18* (4), 303-322.

Suyanto, W. (1998). *The effects of Student Teams-Achievement Divisions on mathematics achievement in Yogyakarta rural primary schools.* Unpublished doctoral dissertation, University of Houston.

Tarver, S. G. & Jung, J. A. (1995). A comparison of mathematics achievement and mathematics attitudes of first and second graders instructed with either a discovery-learning mathematics curriculum or a Direct Instruction curriculum. *Effective School Practices, 14* (1), 49-57.

Teelucksingh, E., Ysseldyke, J., Spicuzza, R., & Ginsburg-Block, M. (2001). *Enhancing the learning of English language learners: Consultation and a curriculum based monitoring system*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted, 29* (1), 60-89.

Tingey, B. and Simon, C. (2001). *Relationship study for SuccessMaker levels and SAT-9 in Hueneme 3 Elementary District, school year 2000-2001, with growth analysis.* Pt. Hueneme, CA (Pearson Education Technologies 2001).

Tingey, B. & Thrall, A. (2000). *Duval County Public Schools evaluation report for 1999-2000.* Duval County, FL, (Pearson Education Technologies 2000).

Thomas, J.A. (1999). The effects of computer-based instruction on the academic achievement of at-risk students.

Tsuei, M. (2005, June). *Effects of a web-based curriculum-based measurement system on math achievement: A dynamic classwide growth model*. Paper presented at the National Educational Computing Conference, Philadelphia, PA.

Turner, L.G. (1985). *An evaluation of the effects of paired learning in a mathematics computer-assisted-instruction program*. Unpublished doctoral dissertation, Arizona State University.

Tusher, L. (1998). *A three-year longitudinal study assessing the impact of the SuccessMaker program on student achievement and student and teacher attitudes in the methacton School District elementary schools*. Lehigh, PA: Lehigh University.

Underwood, J., Cavendish, S., Dowling, S., Fogelman, K., & Lawson, T. (1996). Are integrated learning systems effective learning support tools? *Computers Education, 26* (1-3), 33-40.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Vanderheyden, A.M. & Burns, M.K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30* (3), 15-31.

Vaughan, W. (2002). Effects of cooperative learning on achievement and attitude among students of color. The Journal of Educational Research, 95 (6), 359-364.

Villasenor, A., & Kepner, H.S. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24* (1), 62-69.

Vogel, J.J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C.A. (2006). Using Virtual Reality with and without Gaming Attributes for Academic Achievement. *Journal of Research on Technology in Education*, 39(1), 105-118.

Vreeland, M., Vail, J., Bradley, L., Buetow, C., Cipriano, K., Green, C., Henshaw, P., & Huth, E. (1994). Accelerating cognitive growth: The Edison school math project. *Effective School Practices, 13* (2), 64-69.

Wagreich, P., Goldberg, H., & TIMS Project Staff. (1997). *Math Trailblazers: A mathematical journey using science and language arts.* Dubuque, IA: Kendall/Hunt.

Waite, R.E. (2000). *A study on the effects of Everyday Mathematics on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district*. Unpublished doctoral dissertation, University of North Texas.

Watkins, M.W. (1986). Microcomputer-based math instruction with first-grade students. *Computers in Human Behavior, 2,* 71-75.

Waxman, H. C., & Huang, S. L. (1996-97). Differences by level of technology use on students' motivation, anxiety, and classroom learning environment in mathematics. *Journal of Educational Technology Systems, 25*(1), 67-77.

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. (1995). *The evaluation of Project SEED, 1991-94: Detroit Public Schools*. Detroit, MI: Detroit Public Schools.

Webster, A.H. (1990). *The relationship of computer assisted instruction to mathematics achievement, student cognitive styles, and student and teacher attitudes*. Unpublished doctoral dissertation, Delta State University, Cleveland, MS.

Webster, W., & Chadbourn, R. (1996). *Evaluation of Project SEED through 1995*. Dallas, TX: Dallas Public Schools.

Webster, W.J. (1992). *The evaluation of Project SEED: 1991-1992: Detroit Public Schools.* Detroit, MI: Detroit Public Schools.

Webster, W.J. (1999*). The national evaluation of Project SEED in five school districts 1997-1998*. Dallas, TX: Webster Consulting.

Webster, W.J., & Chadbourn, R.A. (1989). *Final report: The longitudinal effects of SEED instruction on mathematics achievement and attitudes*. Dallas, TX: Dallas ISD.

Webster, W.J., & Chadbourn, R.A. (1992). *The evaluation of Project SEED: 1990-91*. Dallas, TX: Dallas ISD.

Webster, W.J., & Chadbourn, R.A. (1992). *The evaluation of Project SEED: 1989-90*. Dallas, TX: Dallas ISD.

Webster, W.J., & Dryden, M. (1998). *The evaluation of Project SEED: 1997-1998: Detroit Public Schools.* Detroit, MI: Detroit Public Schools.

Weidinger, D. (2006). *The effects of classwide peer tutoring on the acquisition of kindergarten reading and math skills.* Unpublished doctoral dissertation, University of Kansas.

Wellington, J. (1994). Evaluating a mathematics program for adoption: Connecting Math Concepts. *Effective School Practices, 13* (2), 70-75.

Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics.* Princeton, NJ: Educational Testing Service Policy Information Center.

What Works Clearinghouse (2005). *Middle school math curricula*. Washington, DC: U.S. Department of Education. Retrieved July 6, 2006, from http://www.whatworks.ed.gov/Topic.asp?tid=03&ReturnPage=default.asp

What Works Clearinghouse (2006). *Elementary school math.* Washington, DC: U.S. Department of Education. Retrieved September 21, 2006, from http://w-w-c.org/Topic.asp?tid=04&ReturnPage=default.asp

Whitaker, J.C. (2005). *Impact of an integrated learning system on reading and mathematics achievement*. Unpublished doctoral dissertation, Tennessee State University.

White, M.S. (1996). *The effects of TIPS Math activities on fourth grade student achievement in mathematics and on primary caregivers' and teachers' opinions about primary caregivers' involvement in schoolwork.* Unpublished master's thesis, University of Maryland.

Williams, D. (2005*). The impact of cooperative learning in comparison to traditional instruction on the understanding of multiplication in third grade students.* Unpublished doctoral dissertation, Capella University.

Wilson, C. L., & Sindelar, P.T. (1991). Direct Instruction in math word problems: Students with learning disabilities. *Exceptional Children*, *57*(6), 512-519.

Wisconsin Center for Education Research. (2000). Cognitively Guided Instruction & systemic reform. *The newsletter of the Comprehensive Center-Region VI, 5 (2).* Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Wodarz, N. (1994). *The effects of computer usage on elementary students' attitudes, motivation and achievement in mathematics.* Unpublished doctoral dissertation, Northern Arizona University.

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings. *Exceptional Children, 63* (3), 373–389.

Xin, J.F. (1999). Computer-assisted cooperative learning in integrated classrooms for students with and without disabilities. *Information Technology in Childhood Education*, *1* (1), 61-78.

Yager, S., Johnson, R. T., Johnson D. W., & Snider, B. (1986). The impact of group processing on achievement in cooperative learning groups. *The Journal of Social Psychology*, *126* (3), 389-397.

Ysseldyke, J., Betts, J. Thill, T., & Hannigan, E. (2004). Use of an instructional management system to improve mathematics skills for students in Title I programs. *Preventing School Failure, 48* (4), 10-17.

Ysseldyke, J.E. & Bolt, D. (2006). *Effect of technology-enhanced progress monitoring on math achievement.* Minneapolis, MN: University of Minnesota.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96* (3), 163-173.

Ysseldyke, J. Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8* (20), 247-265.

Ysseldyke, J. Tardew, S., Betts, J., Thill, Teri, & Hannigan, E. (2004). Use of an instructional management system to enhance math instruction of gifted and talented students. *Journal of Education for the Gifted, 27* (4), 293-310.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Ysseldyke, J., & Tardrew, S. (2002). *Differentiating math instruction: A large-scale study of Accelerated Math (Final report).* Wisconsin Rapids, WI: Renaissance Learning, Inc.

Ysseldyke, J., & Tardrew, S. (2005). *Use of a progress monitoring system to enable teachers to differentiate math instruction*. Minneapolis, MN: University of Minnesota.

Zollman, A., Oldham, B., & Wyrick, J. (1989). *Effects of computer-assisted instruction on reading and mathematics achievement of Chapter 1 students*. Lexington, KY: Fayette County Public Schools.

Zuber, R.L. (1992). *Cooperative learning by fifth-grade students: The effects of scripted and unscripted techniques.* Unpublished doctoral dissertation, Rutgers University.

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

TABLE 1									
Mathematics Curricula: Descriptive Information and Effect Sizes for Qualifying Studies									
Study	**Design**	**Duration**	**N**	**Grade**	**Sample Characteristics**	**Evidence of Initial Equality**	**Effect Sizes by Posttest and Subgroup**	**Effect Sizes**	**Overall Effect Size**
Everyday Mathematics, Math Trailblazers, and Investigations in Number, Data, and Space (ARC Study)									
Sconiers, Isaacs, Higgins, McBride, & Kelso (2003)	Matched Post Hoc	1 year	742 schools 100,875 students	3-5	Schools across Illinois, Massachusetts, and Washington	Separate matching routines were carried out for each of the five state-grade combinations. Schools on publishers' lists were matched with control schools on variables considered to be best predicators of math achievement - mainly reading scores, SES, and ethnicity	**ISAT/MCAS/ ITBS/WASL**		+0.10
							Computation	+0.10	
							Measurement	+0.14	
							Geometry	+0.08	
							Prob/Stats	+0.03	
							Algebra	+0.09	
							Race/ Ethnicity		
							Asian	+0.11	
							Black	+0.09	
							Hispanic	+0.02	
							White	+0.10	
							SES		
							Low	+0.11	
							Middle	+0.08	
							High	+0.10	
Everyday Mathematics									
Woodward & Baxter (1997)	Matched	1 year	3 schools 38 students	3	Middle class, suburban schools in Pacific Northwest. Students scoring below 34th percentile	Matched with academically low performing students using Heath Mathematics. Pretests showed no significant differences between the groups	**ITBS**		-0.25
							Computations	-0.22	
							Concepts	+0.10	
							Prob. Solving	-0.10	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

SRA/ McGraw-Hill (2003)	Matched Post Hoc	1 year	562 schools 39,701 students	3-5	Schools across Illinois, Massachusetts, and Washington (subset of Sconiers et al study)	Separate matching routines were carried out for each of the five state-grade combinations. Schools on publishers' lists were matched with control schools on variables considered to be best predicators of math achievement - mainly reading scores, SES, and ethnicity	**ISAT/MCAS/ ITBS/WASL**		+0.12
							Computation	+0.13	
							Measurement	+0.15	
							Geometry	+0.12	
							Probability	+0.04	
							Algebra	+0.07	
							Prob. Solving	+0.05	
							Race/ Ethnicity		
							Asian	+0.11	
							Black	+0.11	
							Hispanic	+0.02	
							White	+0.13	
							SES		
							Low	+0.14	
							Middle	+0.09	
							High	+0.10	
Riordan & Noyce (2001)	Matched Post Hoc	2-4+ years	145 schools 8,793 students	4	Schools across Massachusetts. Mostly white, non-free lunch	Schools on publishers' lists matched with controls on prior state tests, SES	**MCAS**		+0.25
							2-3 years of EM	+0.15	
							4+ years of EM	+0.34	

Waite (2000)	Matched Post Hoc	1 year	Schools: 6 EM, 12 control Students: 732 EM, 2,704 control	3-5	Urban district in northern Texas	Matched with controls on prior mathematics test, SES, and ethnicity	**TAAS**		+0.26
							Operations	+0.25	
							Prob. Solving	+0.31	
							Concepts	+0.24	
							Race/ Ethnicity		
							White	+0.33	
							Black	+0.34	
							Hispanic	0.00	
							SES		
							Low	+0.27	
							Middle-High	+0.18	
							Grade		
							Grade 3	+0.18	
							Grade 4	+0.12	
							Grade 5	+0.29	
Math Trailblazers									
Kelso, Booton, and Majumdar (2003)	Matched Post Hoc	1 year	4,942 students	3,4	Schools across Washington State (subset of Sconiers et al study)	Schools on publishers' lists matched with controls on reading, SES, and ethnicity	**ITBS/WASL**		+0.06
							Computations	0.00	
							Concepts	+0.09	
							Prob. Solving	+0.07	
Saxon Math									
Resendez & Azin (2005)	Matched Post Hoc	1-5 years	340 schools	1-5	Georgia public schools	Saxon schools matched with the closest non-Saxon site based on SES, race/ethnicity, and other variables	**CRCT**		+0.02

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Scott Foresman-Addison Wesley									
Resendez & Sridharan (2006) and Resendez & Azin (2006)	Random assignment	1 year	4 schools 39 teachers 901 students	3,5	Schools in Ohio and New Jersey	Random assignment was done within the four schools among 3rd and 5th grade teachers separately. Student level pretest scores favored control group. Important demographic characteristics and achievement measures were used as covariates in multilevel models in order to equate the groups	**Terra Nova**	-	-0.01
							Math Total	-0.07	
							Computations	+0.05	
Resendez & Sridharan (2005) and Resendez & Azin Manley (2005)	Random assignment	1 year	6 schools 35 teachers 719 students	2,4	Schools across WY, WA, KY, and VA	Random assignment was done within the six schools among 2nd and 4th grade teachers separately. Student level pretest scores tended to favor the control group	**Terra Nova**	-	+0.04
							Math Total	+0.11	
							Computations	-0.04	
Houghton Mifflin									
Edstar, Inc. (2002)	Matched Post Hoc	1 year	16 districts 297 schools	2-5	Districts across California	HM districts matched to control districts on prior math achievement, student demographic	**SAT-9**		+0.14

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

						variables, and district sizes. The HM and control districts had similar baseline math achievement scores			
Growing with Mathematics									
Biscoe & Harris (2004)	Matched Post Hoc	1 year	144 classrooms	K-5	Schools across Arkansas, Hawaii, Iowa, Oklahoma, and New Jersey	Students were matched as much as possible on grade, race, and math pretest scores	**Terra Nova**		+0.22
							Comprehension	+0.20	
							Computation	+0.23	
Excel Math									
Mahoney (1990)	Matched Post Hoc	1 year	6 schools Students: 221 Excel, 273 control	2,4	Schools in Palm Springs Unified School District, California	Schools were matched on SES, ethnicity, and achievement. Pretreatment achievement differences were adjusted for by using the SAT pretest as a covariate	**SAT**		+0.13
MathSteps									
Chase, Johnston, Delameter, Moore, & Golding (2000)	Matched Post Hoc	1 year	Students: 2,422 treatment, 1,805 control	3-5	Schools across five California school districts	MathSteps schools matched with control schools on school characteristics, student demographics, and past math achievement	**SAT-9**		+0.03
							Grade 3	+0.11	
							Grade 4	+0.03	
							Grade 5	-0.04	
Knowing Mathematics									
Houghton	Matched	12 weeks	4 schools	4-6	Schools in	KM and control	**MAT-8**		+0.10

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Mifflin (n.d)	Post Hoc		39 students		Lincoln, Nebraska	students were well matched on demographics and prior math achievement	Computation	+0.20	
							Problem Solving	+0.14	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

TABLE 2									
Computer-Assisted Instruction: Descriptive Information and Effect Sizes for Qualifying Studies									
Study	**Design**	**Duration**	**N**	**Grade**	**Sample Characteristics**	**Evidence of Initial Equality**	**Effect Sizes by Posttest and Subgroup**	**Effect Sizes**	**Overall Effect Size**
<u>Jostens/Compass Learning</u>									
Alifrangis (1991)	Random assignment	1 year	1 school 12 classes 250 students	4-6	School at an army base near Washington, D.C.	Students were grouped through stratified random assignment to ensure an equal distribution by sex, minority, and ability. Classes were then randomly selected to use the reading or math curriculum - the reading classes served as the control group for the mathematic classes	**CTBS**		-0.08
Becker (1994)	Random assignment	1 year	1 school 8 classes	2-5	Inner city east-coast school	Groups of equal prior CAT scores were assigned to computer math or computer reading - these "half classes" served as controls for each other	**CAT**		+0.07
							Computations	+0.10	
							Concepts and Applications	-0.02	
Zollman, Oldham, & Wyrick (1989)	Matched	1 year	15 schools Students: 146 treatment 274 control	4-6	Lexington, Kentucky. Chapter 1 students	Students matched on Chapter 1 status	**MAT6**		+0.21
Hunter (1994)	Matched	1 year	6 schools 150 students	2-5	Public schools in Jefferson County, Georgia. Chapter 1 students	Treatment groups were well matched to control groups on ethnicity and SES. ANCOVA used to adjust posttest scores for pretest differences	**ITBS**		+0.40

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Estep, McInerney, Vockell, & Kosmoski (1999-2000)	Matched Post Hoc	1-5 years	106 schools	3	Schools across Indiana	Jostens schools were well matched with control schools on Indiana's ISTEP	**ISTEP**		+0.01
							Computation	-0.03	
							Prob. Solving	0.00	
Spencer (1999)	Matched Post Hoc	5 years	92 students	2,3	Urban school district in southeastern Michigan	Experimental students were well matched to control students on gender, race, and past CAT total math scores	**CAT**		+0.40
							Grade 2 Starters	+0.37	
							Grade 3 Starters	+0.44	
Clariana (1994)	Matched Post Hoc	1 year	1 school 4 classes 85 students	3	School in a predominantly white, rural area	Classes were taught by same teacher in successive years - math scores from the California Test of Basic Skills were not statistically different but favored the control group. Differences were accounted for in final analyses	**CTBS**		+0.66
<u>CCC/Successmaker</u>									
Ragosta (1983)	Random assignment	3 years	4 schools	1-6	Schools in the Los Angeles Unified School District	Alternate waves of students received or didn't receive CAI over the years. All students were pretested with the ITBS and these scores, sex, ethnicity, and classroom differences were accounted for in final analyses	**CTBS**		+0.36
							Computations	+0.72	
							Concepts	+0.09	
							Applications	+0.26	
Hotard & Cortez (1983)	Random assignment	6 months	2 schools 190 students	3-6	Schools in Lafayette Parish, Louisiana	Students were matched on past math CTBS scores and then each member of the matched pair was randomly assigned	**CTBS**		+0.19
Manuel (1987)	Random assigment	12 weeks	3 schools 165 students	3-6	Schools in Omaha, Nebraska	Students were randomly assigned to CAI or control conditions, stratifying on ability and pretest	**CTBS**		+0.07

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Mintz (2000)	Matched Post Hoc	1 year	8 schools 487 students	4,5	Schools in Etowah County, Alabama	Schools were matched on similar SAI mean scores and students were well matched on IQ and achievement scores	**SAT-9**		-0.06
Laub (1995)	Matched Post Hoc	5 months	2 schools 14 classes 314 students	4,5	Schools in Lancaster County, Pennsylvania	Treatment students were matched with controls (previous 4th/5th grade classes) on SAT Total Math - ANCOVA was used to adjust initial group differences	**SAT**		+0.27
Classworks									
Patterson (2005)	Random assignment	14 weeks	30 students	3	Rural school in central Texas	The two groups of students were nearly identical at pretest	**SAT-9**		+0.85
Whitaker (2005)	Matched Post Hoc	1 year	2 schools 218 students	4,5	Schools in rural Tennessee	The two schools were well matched on demographics and TCAP pretests	**TCAP**		+0.21
Lightspan									
Birch (2002)	Matched Post Hoc	2 years	2 schools 101 students	2,3	Schools in the Caesar Rodney School District in Delaware	The treatment and control schools used the same district-wide curriculum, had access to comparable resources and financial resources, and had similar demographics. Any pretreatment differences were adjusted for by using the pretest as a covariate	**SAT**		+0.28
							end of year 1	+0.53	
							end of year 2	+0.28	
Other CAI									
Becker (1994) (CNS)	Random assignment	1 year	1 school 9 classes	2-5	Inner city east-coast school	Groups of equal prior CAT scores were assigned to computer math or computer reading - these "half classes" served as controls for each other	**CAT**		+0.18
							Computations	+0.18	
							Concepts and Applications	+0.12	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Carrier, Post, & Heck (1985) (various CAI)	Random assignment	14 weeks	6 classes 144 students	4	Metropolitan school district in Minnesota	Teachers ranked their students from high to low on the basis of ITBS math scores. Students were then paired so that the students with ranks 1 and 2 formed the first pair; 3 and 4, the second pair; and so on. They randomly assigned one member of each pair to the computer treatment and the other to the worksheet treatment	**Experimenter-designed Test**		+0.21
							Symbolic Algorithms	+0.01	
							Division Facts	+0.28	
							Multiplication Facts	+0.35	
Abram (1984) (The Math Machine)	Random assignment	12 weeks	1 school 5 classes 103 students	1	Suburban school district in Southwest	Half the males and females scoring in each quartile of the ability level test were randomly assigned to either receive computer phonics or computer math. Each group served as the control for the other	**ITBS**		-0.18
Watkins (1986) (The Math Machine)	Random assignment	6 months	1 school 82 students	1	Suburban southwestern school	ITBS and Cognitive Abilities Test served as covariates	**CAT**		+0.41
Fletcher, Hawley, & Piele (1990) (Milliken Math Sequences)	Random assignment	4 months	1 school 4 classes 79 students	3,5	School in rural Saskatchewan	Treatment groups were well matched at pretest. Posttests adjusted for pretests	**Canadian Test of Basic Skills**		+0.40
							Grade 3	+0.48	
							Computations	+0.58	
							Concepts	+0.20	
							Problem Solving	+0.10	
							Grade 5	+0.32	
							Computations	+0.22	
							Concepts	+0.30	
							Problem Solving	+0.36	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Van Dusen & Worthen (1994) (unspecified program)	Randomized quasi-experiment	1 year	6 schools 141 classes 4,612 students	K-6	Schools selected from diverse geographic areas across the U.S.	Classes within each school were randomly assigned to one of three implementation conditions. Separate analyses of covariance, using previous year's NRT scores, were conducted in math and reading to equate all three groups on achievement	**Norm-Referenced Tests**		+0.01
							Good Implementers	+0.05	
							Weak Implementers	-0.04	
Shanoski (1986) (Mathematics Courseware)	Randomized quasi-experiment	20 weeks	4 schools 32 classes 832 students	2-6	Rural Pennsylvania	Students were well matched on the math portion of the CAT	**CAT**		-0.02
Turner (1985) (Milliken Math Sequencing and Pet Professor)	Randomized quasi-experiment	15 weeks	275 students	3-4	School in suburb of Phoenix, Arizona	The CAI and control groups did not differ with respect to CTBS pretest scores	**CTBS**		+0.37
Schmidt (1991) (Wasatch ILS)	Matched	1 year	4 schools 1,224 students	2-6	Schools in Southern California	Schools were matched on SES and CTBS scores. Students were split into high- and low-achievement standings according to CTBS scores, providing a measure of control for initial differences	**CTBS**		+0.05
							Low Achievers	+0.07	
							High Achievers	+0.03	
Bass, Ries, & Sharpe (1986) (CICERO software)	Matched	1 year	1 school 178 students	5-6	School in rural Virginia. Chapter 1 students	Two groups were comparable at pretest. Pretest scores were used as covariates	**SRA Achievement Series**		+0.02

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Webster (1990) (Courses by Computers Math)	Matched	14 weeks	5 schools 120 students	5	Schools in rural Missipi Delta school district	Students from the experimental and control groups shared similar demographics. There were non-significant differences in math pretest scores favoring the control group. Posttests adjusted for pretests	**SAT**		+0.13
Hess & McGarvey (1987) (Memory, Number Farm)	Matched	5 months	66 students	K	Schools drew students from wide range socio-economic backgrounds	Each student in the home-use class was matched on pretest readiness score with a student in the classroom-use group and one in the control group. Students were also matched by gender when possible	**Criterion-Referenced Test**		+0.14
Gilman & Brantley (1988)	Matched	1 year	1 school 57 students	4	School in rural Indiana	Two groups were comparable at pretest. Pretest scores were used as covariates	**ITBS**		+0.03
Miller (1997) (Waterford Integrated Learning System)	Matched Post Hoc	1 to 3 years	Schools: 10 WILS, 20 control	3-5	New York City Public Schools	Two comparison schools were matched to each treatment school based on student achievement, SES, race/ethnicity, LEP, and attendance	**MAT**		+0.17
Levy (1985) (Mathematics Strands, Problem Solving - ISI)	Matched Post Hoc	1 year	4 schools 576 students	5	Suburban New York School District	Students from the two cohorts shared similar demographics, teachers, and curriculum. There were no significant differences between cohorts on math pretest scores	**SAT**		+0.21
Schreiber, Lomis, & Nys (1988) (WICAT)	Matched Post Hoc	3 years	Schools: 1 WICAT, 3 control 254	1-4	Schools in Dearborn, Michigan	WICAT school matched to control schools on ethnicity and Cognitive Abilities Test	**ITBS**		+0.48

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

			students						
Stone (1996) (Exploring Measurement Time and Money)	Matched Post Hoc	3 years	2 schools Students: 40 CAI, 74 control	2	Middle-class schools	Students were matched on SES and end-of-first grade cognitive abilities tests, which were used as covariates in analyses of covariance	**ITBS**		+1.16
Barton (1988)	Matched Post Hoc	1 year	1 school Students: 36 CAI, 56 control	5	Suburban school near San Diego	CAI Students were compared to the two previous 5th grade cohorts - they shared shared similar demographics, teachers, and curriculum. There were no significant differences between cohorts on math pretest scores	**CTBS**		+0.68
<u>Computer-Managed Learning Systems</u>									
<u>Accelerated Math</u>									
Ysseldyke & Bolt (2006)	Randomized quasi-experiment	1 year	5 schools 823 students	2-5	Schools in Texas, Alabama, South Carolina, and Florida	Teachers were randomly assigned to AM or control. Classes were well matched on demographics and achievement. Pretest scores and schools were used as covariates	**Terra Nova**		+0.03
Ross & Nunnery (2005)	Matched Post Hoc	1 year	2350 AM students, 1841 control students	3-5	Schools in southern Mississippi	A matched control school was selected for each School Renaissance school on the basis of student ethnicity, poverty, mobility, school location, grades served, size, prior achievement (school means on 2000-01 MCT reading and math), and no or very limited usage of Accelerated Reader or	**MCT**		+0.04

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

						Accelerated Math			

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, Boys, & Lemkuil, (2003)	Matched Post Hoc	1 year	Students: 397 AM, 913 control	3-5	Schools in large urban district in the midwest	The first control group consisted of 484 students from same schools as students who participated in AM. The second consisted of 429 students randomly selected from rest of district from annual testing database. Analysis of covariance was performed with pretest scores as covariate	**NALT**		+0.11
							within class comparison	+0.08	
							district comparison	+0.14	
Spicuzza, Ysseldyke, Lemkuil, Kosciolek, Boys, & Teelucksingh (2001)	Matched Post Hoc	5 months	Students: 137 AM, 358 control	4,5	Large urban district in the midwest	The first control group consisted of 61 students from the same schools as students who participated in AM. The second consisted of 297 students selected from the district. AM schools were matched with control schools on SES, ELL status, and other demographics. Analysis of covariance was performed with pretest scores as covariate	**NALT**		+0.17
							within class comparison	+0.19	
							district comparison	+0.14	
Johnson-Scott (2006)	Matched Post Hoc	1 year	3 schools 7 classes 82 students	5	Schools in rural Mississippi	AM students were well matched to control students on race/ethnicity, SES, and achievement scores.	**MCT**		+0.23

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

TABLE 3

Instructional Process Strategies: Descriptive Information and Effect Sizes for Qualifying Studies

Study	Design	Duration	N	Grade	Sample Characteristics	Evidence of Initial Equality	Effect Sizes by Posttest and Subgroup	Effect Sizes	Overall Effect Size
Cooperative Learning									
Classwide Peer Tutoring									
Greenwood, Delquadri, & Hall (1989)	Randomized quasi-experiment	4 years	123 students	1-4	High poverty school district in metropolitan area of Kansas City, Kansas	IQ scores and SES were not significantly different between the groups. IQ and pretest achievement served as covariates	**MAT**		+0.33
							2 year followup	+0.23	
Peer-Assisted Learning Strategies									
Fuchs, Fuchs, Yazdian, & Powell (2002)	Randomized quasi-experiment	16 weeks	20 classes 323 students	1	Schools in southeastern city	Teacher demographics were comparable across PALS and no-PALS groups. Initial math achievement status also comparable	**SAT**		+0.10
							High Achieving	+0.09	
							Avg. Achieving	+0.10	
							Low Achieving	+0.14	
Fuchs, Fuchs, & Karns (2001)	Randomized quasi-experiment	15 weeks	20 teachers 228 students	K	Schools in southeastern city	There were no statistically significant differences in teacher demographics. SESAT pretest scores were comparable across groups	**SESAT**		+0.24
							High Achieving	-0.41	
							Avg. Achieving	+0.52	
							Low Achieving	+0.51	
							Disability	+0.65	
							SAT		
							High Achieving	+0.85	
							Avg Achieving	-0.20	
							Low Achieving	+0.47	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

							Disability	+0.20	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Student Teams-Achievement Divisions									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school, middle-class students	Treatment and control classes were well matched on SES, class size, age, and pretest scores. Pretest scores served as covariates	**Objective-Based Test**		+0.19
							Computation	+0.36	
							Comprehension	+0.01	
Glassman (1989)	Randomized quasi-experiment	6 months	2 schools 24 classes 441 students	3-5	Schools in diverse, suburban district in Long Island, New York	Classes were randomly assigned after being stratified and paired to use STAD or traditional strategies. Pretest scores served as covariates	**ITBS**		+0.01
Mevarech (1991)	Randomized quasi-experiment	12 weeks	54 students	3	Low SES school in Israel	Initial achievement indicated no significant difference between groups prior to the beginning of the study	**Teacher-Designed Test**		+0.60
							Low Achievers	+0.86	
							High Achievers	+0.69	
Suyanto (1998)	Matched	4 months	10 schools 30 classes 664 students	3-5	Schools across rural Indonesia	STAD and control classes were well matched on demographics and pretest scores. Pretest scores served as covariates	**Indonesian Elementary School Test of Learning**		+0.40
Student Team Mastery Learning									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school, middle-class students	Treatment and control classes were well matched on SES, class size, age, and	**Objective-Based Test**		+0.29
							Computation	+0.36	

							pretest scores. Pretest scores served as covariates	Comprehension	+0.21	
Mevarech (1991)	Randomized quasi experiment	12 weeks	54 students	3	Low SES school in Israel	Initial achievement indicated no significant difference between groups prior to the beginning of the study	**Teacher-Designed Test**		+0.55
							Low achievers	+0.80	
							High Achievers	+0.80	
Cooperative/Individualized Programs									
<u>TAI</u>									
Slavin & Karweit (1985)	Random assignment	16 weeks	17 classes 382 students	3-5	Hagerstown, Maryland	Pretests non-significantly favored TAI group in comparison to MMP and untreated control. Differences controlled for in analyses	**CTBS**		+0.28
							TAI vs MMP		
							Computations	+0.39	
							Concepts/ Applications	+0.01	
							TAI vs Control		
							Computations	+0.67	
							Concepts/ Applications	+0.06	
Slavin & Karweit (1985)	Random assignment	18 weeks	10 classes 212 students	4-6	Inner-city, Wilmington, Delaware	Students were well matched on pretest, although TAI scored non-significantly higher than MMP. Differences controlled for in analyses	**CTBS**		+0.38
							Computations	+0.76	
							Concepts and Applications	0.00	
Slavin, Madden, & Leavey (1984a)	Matched	24 weeks	59 classes 1,367 students	3-5	Schools located in middle class suburb of Baltimore, Maryland.	TAI and control schools were well matched on CAT pretests	**CTBS**		+0.14
							Computations	+0.18	
							Concepts and Applications	+0.10	
							Students with special needs		

							Computations	+0.19	
							Concepts and Applications	+0.23	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Stevens & Slavin (1995)	Matched	2 years	45 classes 873 students	2-6	Schools located in diverse Baltimore suburb.	Treatment classes were matched with control classes on CAT scores, ethnicity, and SES. There were significant pretest differences on Total Math scores that favored the comparison students. Differences were controlled for in analyses	**CAT**		+0.20
							Computations	+0.29	
							Applications	+0.10	
							Students with special needs		
							Computations	+0.59	
							Applications	+0.35	
							Gifted Students		
							Computations	+0.59	
							Applications	+0.19	
Karper & Melnick (1993)	Matched	1 year	165 Students	4,5	School in affluent suburb of Harrisburg, Pennsylvania	TAI and control classes were well matched on district test	**District Standardized Test**		-.0.11
							Computation	-0.11	
							Concepts	-0.22	
<u>Project CHILD</u>									
Orr (1991)	Matched	3 years	2 schools 186 students	2-5	Schools in Northeast and Northwest Florida	Project CHILD was used by a subset of students in each school and they were well matched on standardized test scores to students within the same school. Prestest differences were controlled for in final analyses	**Standardized Achievement Tests**		+0.69

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Direct Instruction									
<u>Connecting Math Concepts</u>									
Snider & Crawford (1996)	Random assignment	1 year	1 school 46 students	4	Rural school in northern Wisconsin	There were no significant pretest differences between the CMC classroom and Invitations to Mathematics (Scott Foresman)	**NAT**		+0.26
							Computation	+0.72	
							Concepts/ Prob.Solving	+0.01	
Tarver & Jung (1995)	Matched	2 years	88 students: 21 CMC, 67 MTW/CGI	1,2	School in suburban midwest	Control students in Mathematics Their Way combined with CGI classes scored higher on pretest but the difference was not significant. Differences were accounted for in final analyses	**CTBS**		+1.33
							Computations	+2.13	
							Concepts/Apps	+0.68	
Crawford & Snider (2000)	Matched Post Hoc	1 year	1 school 52 students	4	Rural school in northern Wisconsin	CMC students were compared to students from the same teacher's class the previous year. The classes were nearly identical on NAT pretest scores	**NAT**		+0.41
							Computations	+0.33	
							Concepts/ Prob.Solving	+0.39	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

User-Friendly Direct Instruction									
Grossen & Ewing (1994)	Randomized	2 years	1 school 45 students	5,6	School in Boise, Idaho	Students were stratified based on high, medium, and low school performance and then randomly assigned to 2 equivalent groups for treatment. Posttest measures were adjusted for any pretest differences	**Woodcock-Johnson**		+0.52
							Applications	+0.50	
							ITBS		
							Concepts	+0.44	
							Prob. Solving	+0.17	
							Operations	+0.97	
Mastery Learning									
Mastery Learning									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school - middle-class students	Treatment and control classes were well matched on SES, class size, age, and pretest scores. Pretest scores served as covariates	**Objective-Based Test**		+0.42
							Computation	+0.55	
							Comprehension	+0.28	
Mevarech (1991)	Randomized quasi experiment	12 weeks	85 students	3	Low SES school in Israel	Initial achievement indicated no significant difference between groups prior to the beginning of the study	**Teacher-Designed Test**		+1.08
							Low Achievers	+1.10	
							High Achievers	+2.20	
Cox (1986)	Matched	6 months	173 students	5	School in southwestern Missouri	Students were assigned by alternating down a ranked list to mastery learning or control conditions. Posttests were adjusted for pretest achievement differences	**ITBS**		+0.22

Monger (1989)	Matched Post Hoc	1 year	140 students	2,5	Schools located in suburban districts in Oklahoma	Schools shared similar demographics and the two groups of students were fairly well matched on math pretests.	**MAT-6**		-0.18
							Grade 2	-0.09	
							Grade 3	-0.27	
Anderson, Scott, & Hutlock (1976)	Matched Post Hoc	1 year	2 schools	1-6	Schools near Cleveland, Ohio	Students were matched on Metropolitan Readiness Test in grades 1-3, or the Otis-Lennon Intelligence Test in grades 4-6	**CAT**		+0.04
							Computation	+0.17	
							Prob. Solving	+0.07	
							Concepts	-0.12	
Professional Development Focused on Math Content									
Cognitively Guided Instruction									
Carpenter, Fennema, Peterson, Chiang, & Loef (1989)	Randomized quasi-experiment	1 year	40 teachers	1	Schools in and around Madison, Wisconsin	Analyses of covariance between groups were computed on each student achievement measure controlling for prior math achievement	**ITBS**		+0.24
							Computation	+0.22	
							Prob. Solving	+0.25	
Dynamic Pedagogy									
Armour-Thomas, Walker, Dixon-Roman, Mejia, & Gordon (2006)	Matched	1 year	2 schools 120 students	3	Schools in New York suburb	Treatment students were well matched to non-participating students within the same two schools on race/ethnicity, SES, and achievement scores	**Terra Nova**		+0.32

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Professional Development Focused on Classroom Management and Motivation									
Missouri Mathematics Project									
Slavin & Karweit (1985)	Random assignment	16 weeks	22 classes 366 students	3-5	Schools in and around Hagerstown, Maryland	Scores adjusted for CAT (pre) scores separately for each grade	**CTBS**		+0.57
							MMP two-group	+0.84	
							MMP whole-class	+0.29	
Good & Grouws (1979)	Randomized quasi-experiment	3 months	27 schools 40 teachers	4	Tulsa, Oklahoma school district	Schools used the same textbook and were similar in SES. The treatment group began the project with lower achievement scores than the control group. Pretests were used as covariates	**SRA**		+0.33
Ebmeier & Good (1979)	Randomized quasi-experiment	15 weeks	28 schools 39 teachers	4	Schools in large southwestern school district	Schools served predominantly low-SES students. All teachers had taught for at least three years. Pretests were used as covariates	**SRA**		+0.42
Consistency Management & Cooperative Discipline									
Freiberg, Prokosch, Treiser, & Stein (1990)	Matched Post Hoc	1 to 2 years	10 schools 699 students	2-5	Low-performing, high minority schools in Houston, Texas	CM schools were matched to comparison schools on demographics and achievement. Students in CM classrooms were compared to randomly selected (according to grade distribution) students	**MAT6, TEAMS**		+0.40

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

						from non-program schools			
Freiberg, Connell, & Lorentz (2001)	Matched Post Hoc	1 year	7 schools 543 students	4-6	Chapter 1 schools in large urban city in the southwest U.S.. Students mainly Latino	The schools were similar in demographics and pretest scores. Pretest TAAS scores used as a covariate	**TAAS**		+0.33
Opuni (2006)	Matched Post Hoc	1 year	456 students	3	Newark Public Schools	CMCD schools were matched to comparison schools on demographics and achievement. Students in CM classrooms were compared to randomly selected (according to grade distribution) students from non-program schools	**Stanford-9**		+0.53
Supplemental Programs									
<u>Small Group Tutoring</u>									
Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett (2005)	Random assignment	16 weeks	127 students	1	Schools in southeastern metropolitan district. Students at-risk for development of mathematics difficulty	Tutoring and non-tutoring conditions were well matched on demographics, math test scores, and ability	**Woodcock-Johnson** Calculations Applied Prob. **Experimenter-made measure** Addition Subtraction Concepts and Applications	 +0.56 +0.09 +0.36 +0.03 +0.62	+0.37

							Story Problems	+0.59	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Every Day Counts									
RMC Research Corporation (2006)	Matched	6 months	13 schools 587 students	5	Schools in New Haven, CT	Initially, 12 schools were put into low, middle, and higher achievement groups based on CMT math scores. School pairs were then mached on demographics and randomly assigned. Later, 2 schools that had piloted EDC were added to the study. Pretest differences were controlled for in the final analyses	Assessment modeled after **CMT**		+0.15
Project SEED									
Webster (1994)	Matched Post Hoc	1-3 semesters	Students: Cycle 1: 732 Cycle 2: 558 Cycle 3: 249	4-6	Detroit Public Schools	Schools were matched on SES, math scores, and size of school. Each SEED student was matched to a control student on sex, ethnicity, grade, SES, and achievement. Analysis of covariance was used where significant differences were found	**CAT**	-	+0.73
							1 semester SEED exposure		
							Math Total	+0.29	
							Computations	+0.35	
							Concepts	+0.32	
							2 semesters SEED exposure		
							Math Total	+0.60	
							Computations	+0.58	
							Concepts	+0.68	
							3 semesters SEED exposure		
							Math Total	+0.73	

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

							Computations	+0.74	
							Concepts	+0.62	
Webster & Dryden (1998)	Matched Post Hoc	14+ weeks	604 students	3	Detroit Public Schools	Each student in each of the SEED groups was systematically matched to a comparison student by SES, ethnicity, and math pretest scores. Math pretest scores were used as covariates	**MAT-6**		+0.25
							Math Total	+0.25	
							Concepts/ Prob. Solving	+0.28	
							Procedures	+0.17	

Table 4

Summary of Evidence Supporting Currently Available Elementary Mathematics Programs

● Strong Evidence of Effectiveness
Classwide Peer Tutoring (IP)
Missouri Mathematics Program (IP)
Peer Assisted Learning Strategies (PALS) (IP)
Student Teams-Achievement Divisions (IP)
TAI Math (IP/MC)

◒ Moderate Evidence of Effectiveness
Classworks (CAI)
Cognitively Guided Instruction (IP)
Connecting Math Concepts (IP/MC)
Consistency Management & Cooperative Discipline (IP)
Project SEED (IP)
Small-Group Tutoring (IP)

◔ Limited Evidence of Effectiveness
Accelerated Math (CAI)
Dynamic Pedagogy (IP)
Every Day Counts (IP)
Excel Math (MC)
Everyday Mathematics (MC)
Growing with Mathematics (MC)
Houghton-Mifflin Mathematics (MC)
Knowing Mathematics (MC)
Mastery Learning (IP)
Lightspan (CAI)
Project CHILD (IP/CAI)

○ Insufficient Evidence
Math Steps (MC)
Math Trailblazers (MC)
Saxon Math (MC)
Scott Foresman-Addison Wesley (MC)

N No Qualifying Studies
Adventures of Jasper Woodbury (IP/CAI)
AIMSweb® Pro Math (CAI)
Bridges in Mathematics (MC)
Compass Learning (CAI) (Current version)
Corrective Math (MC)

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Count, Notice, & Remember (IP)
Destination Math Series (CAI)
First in Math® (CAI)
Great Explorations in Math and Science (IP/MC)
Harcourt Math (MC)
Investigations in Number, Data, and Space (MC)
Larson's Elementary Math
Math Advantage (MC)
MathAmigo (CAI)
Math Blasters (CAI)
Math Central (MC)
Math Coach (MC/IP)
Math Expressions (MC)
Math Explorations and Applications (MC)
Math in My World (MC)
Math Made Easy (CAI)
Math Matters (IP)
Math Their Way (MC)
Math & Me Series (MC)
Math & Music (CAI)
Mathematics Plus (MC)
Mathematics Their Way (MC)
Mathletics (MC)
MathRealm (CAI)
MathWings (IP/MC)
Macmillan McGraw-Hill Math (MC)
McGraw-Hill Mathematics (MC)
Number Power (MC)
Problem Solving Step by Step (IP/MC)
Progress in Mathematics (MC)
Project IMPACT (IP)
Project M3: Mentoring Mathematical Minds (MC)
Rational Number Project (MC)
Real Math (MC)
Reciprocal Peer Tutoring (IP)
Scott Foresman Math Around the Clock (IP/MC)
Singapore Math (MC)
SkillsTutor/CornerStone2 (CAI)
SuccessMaker (CAI) (Current version)
TIPS Math (IP)
Voyages (IP/MC)
Waterford Early Math (CAI)
Yearly Progress Pro (CAI)

Note:

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

MC: Mathematics Curricula
CAI: Computer-Assisted Instruction
IP: Instructional Process Strategies

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

APPENDIX 1		
Studies Not Included in the Review		
Author	**Cited by**	**Reason not included/ Comments**
CURRICULA		
Everyday Mathematics		
Briars (2004)		no adequate control group
Briars & Resnick (2000)	NRC	no adequate control group
Carroll (1993)	NRC	no adequate control group
Carroll (1994-1995)	NRC	no adequate control group
Carroll (1995)	NRC	inadequate outcome measure
Carroll (1996)	NRC	no adequate control group
Carroll (1996b).	NRC	inadequate outcome measure
Carroll (1996c)	NRC	no adequate control group
Carroll, (1997)		insufficient match, no pretest
Carroll (1998)	NRC	inadequate outcome measure
Carroll (2001)	NRC	no adequate control group
Carroll (2001b)	NRC	insufficient match, no pretest
Carroll & Fuson (1998)	NRC	no adequate control group
Carroll & Porter (1994)	NRC	insufficient match
Druek, Fuson, Carroll, & Bell (1995)	NRC	no adequate control group
Fuson & Carroll (undated)	NRC	no adequate control group
Fuson & Carroll (1997)	NRC	insufficient match, no pretest
Fuson, Carroll, & Drueck (2000)	NRC	no adequate control group
Mathematics Evaluation Committee (1997)		insufficient match
McCabe (2001)		insufficient match, no pretest
Salvo (2005)		duration < 12 weeks
Math Trailblazers		
Carter, Beissinger, Cirulis, Gartzman, Kelso, and Wagreich (2003)		no adequate comparison groups, pretest differences not accounted for
Lykens (2003)		no adequate control group
Saxon Math		
Bolser & Gilman (2003)		no adequate control group
Calvery, Bell, & Wheeler (1993, November)	NRC	insufficient match
Atkeison-Cherry (2004)		duration < than 12 weeks
Fahsl (2001)		insufficient match, posttest only
Hanson & Greene (2000)	NRC	insufficient information, no adjusting at posttest
Nguyen & Elam (1993)		insufficient information
Nguyen (1994)		insufficient information, no adjusting at posttest
Resendez, Sridiharan, & Azin (2006)		pretest equivalence was not established
Scott Foresman-Addison Wesley		
Gatti (2004)		pretest equivalence not established
Simpson (2001)		no adequate control group

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Houghton Mifflin Mathematics		
EDSTAR, Inc. (2004)		insufficient data
Mehrens & Phillips (1986)		no adequate control group
Sheffield, (2004)		no adequate control group
Sheffield, (2005)		no adequate control group
Investigations in Number, Data, and Space		
Austin Independent School District (2001)	NRC	insufficient match, pretest differences not accounted for
Flowers (1998)		insufficient match and outcome measure
Gatti (2004)		pretest equivalence not documented
Goodrow (1998)	NRC	insufficient match and outcome measure
McCormick (2005)		measure inherent to treatment
Mokros, Berle-Carmen, Rubin, & O'Neill (1996)	NRC	insufficient match and outcome measure
Mokros, Berle-Carmen, Rubin, & Wright, (1994).	NRC	inadequate outcome measure, pretest differences not accounted for
Ross (2003)		no adequate control group
Math Their Way		
Shawkey (1989)		insufficient match
MathWings		
Madden, Slavin, & Simon (1997)		no adequate control group
Madden, Slavin, & Simon (1999)		no adequate control group
Number Power		
Cooperative Mathematics Project (1996)		inadequate outcome measure
Progress in Mathematics		
Beck Evaluation & Testing Associates, Inc (2006)		inadequate control group
Rational Number Project		
Cramer, Post, & delMas (2002)		duration < 12 weeks, inadequate outcome measure
Ross & Chase (1999)		test inherent to measure
Real Math (Explorations and Applications)		
Dilworth and Warren (1980)		inadequate outcome measure, no adequate control group
CAI		
Jostens Learning/Compass Learning		
Brandt & Hutchinson (2005)		no adequate control group
Clariana (1996)	Kulik (SRI)	insufficient information provided

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Jamison (2000)		duration < 12 weeks
Leiker (1993)	Kulik (SRI)	treatment and control used different pretests
Mann, Shakeshaft, Becker, & Kottkamp (1999)		no adequate control group
Moody (1994)		no adequate control group
Rader (1996)		duration < 12 weeks
Roy (1993)	Kulik (SRI)	insufficient information provided
Sinkis (1993)	Kulik (SRI)	insufficient match
Stevens (1991)	Kulik (SRI)	pretest differences too large
Taylor (1990)		no adequate control group
CCC/SuccessMaker		
Crenshaw (1982)		no adequate control group
Donnelly (2004, April)		insufficient match, no adjusting at posttest
Kirk (2003)		no adequate control group
Laub & Wildasin (1998)		no adequate control group
McWhirt, Mentavlos, Rose-Baele, & Donnelly (2003)		no adequate control group
Phillips (2001)		inadequate outcome measure
Tingey & Simon (2001)		no adequate control group
Tingey & Thrall (2000)		no adequate control group
Tusher (1998)		no adequate control group
Underwood, Cavendish, Dowling, Fogelman, & Lawson (1996)	Kulik (SRI)	no evidence of pretest equivalence
Lightspan/Plato		
Giancola (2000)		no adequate control group
Gwaltney (2000)		treatment confounded with other programs
Interactive, Inc. (2001)		program began before pretest
Interactive, Inc. (2002)		program began before pretest
Quinn & Quinn (2001)		no adequate control group
Quinn & Quinn (2001)		no adequate control group
Other CAI		
Axelrod, McGregor, Sherman, & Hamlet (1987)		no adequate control group, duration < 12 weeks
Bedell (1998)		no adequate control group
Brown & Boshamer (2000) (Fundamentally Math)		pretest equivalence not demonstrated
Carrier, Post, & Heck (1985)		inadequate outcome measure
Chang, Sung, and Lin (2006)		duration < 12 weeks
Chiang (1978)		insufficient match
Cognition and Technology Group at Vanderbilt. (1992)		inadequate outcome measure
Dahn (1992) (Wasach)		no evidence of initial equivalence
Emihovich & Miller (1988)		duration < 12 weeks

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Faykus (1993) (WICAT)		duration < 12 weeks
Flaherty Connolly, & Lee-Bayha (2005) (First in Math)		no adequate control group
Haynie (1989)		no adequate control group
Isbell (1993)		no adequate control group
Kastre, Norma Jane		duration < 12 weeks
Lin, Podell, & Tournaki-Rein (1994)		duration < 12 weeks
McDermott & Watkins (1983)		insufficient data
Mevarech & Rich (1985)		no accounting for pretest differences
Mills (1997)		no adequate control group
Orabuchi (1992)		no accounting for pretest differences
Perkins (1987)		duration < 12 weeks
Podell, Tournaki-Rein, & Lin (1992)		duration < 12 weeks
Shiah, Mastropieri, Scruggs, & Fulk (1994-1995)		inadequate outcome measure
Snow (1993)		no adequate control group
Sullivan (1989)		no adequate control group
Suppes, Fletcher, Zanotti, Lorton, & Searle (1973)		no adequate control group
Trautman, T. and Howe, Q.		no adequate control group
Trautman, T.S. and Klemp, R.		no adequate control group
Vogel, Greenwood-Ericksen, Cannon-Bowers, & Bowers (2006)		duration < 12 weeks
Wenglinsky, H. (1998).		no adequate control group
Wodarz (1994)		pretest equivalence not demonstrated
<u>Accelerated Math</u>		
Atkins (2005)		pretest equivalence not established
Boys (2003)		pretest equivalence not established
Brem (2003)		inadequate outcome measure, no adequate control group
Kosciolek (2003)		no adequate control group
Holmes and Brown (2003)		no adequate control group
Leffler (2001)		no adequate control group
Teelucksingh, Ysseldyke, Spicuzza, & Ginsburg-Block (2001)		pretest equivalence not established
Ysseldyke, Spicuzza, Kosciolek, & Boys (2003)		large pretest differences
Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, Boys, & Lemkuil (2003)		insufficient match, pretest differences too large
Ysseldyke & Tardrew (2005)		inadequate outcome measure
Ysseldyke & Tardrew (2002-2003)		inadequate outcome measure
Ysseldyke, Tardrew, Betts, Thill, & Hannigan (2004)		inadequate outcome measure
Ysseldyke, Thill, & Hannigan (2004)		inadequate outcome measure
INSTRUCTIONAL PROCESS STRATEGIES		
<u>Classwide Peer Tutoring</u>		
Greenwood, Dinwiddie, Terry, Wade, Stanley, Thibadeau, & Delquadri (1984)		no adequate control group, inadequate outcome measure

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

DuPaul, Ervin, Hook, & McGoey (1998)		no adequate control group, inadequate outcome measure
Peer-Assisted Learning		
Fuchs, Fuchs, Phillips, Hamlett, & Karns (1995)		test inherent to treatment
Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka (1997)		test inherent to treatment
Student Team-Achievement Divisions		
Vaughan (2002)		no adequate control group
Team Assisted Individualization (TAI)		
Bryant (1981)		duration < 12 weeks
Slavin, Leavey, & Madden (1984)		duration < 12 weeks
Slavin, Madden, & Leavey (1984)		duration < 12 weeks
Project Child		
Butzin (2001)		pretest equivalence not demonstrated
Butzin and King (1992)		pretest equivalence not demonstrated
Florida TaxWatch (2005)		pretest equivalence not demonstrated
Gll (1995)		pretest equivalence not demonstrated
Kromhout (1993)		pretest equivalence not demonstrated
Kromhout & Butzin (1993)		pretest equivalence not demonstrated
Direct Instruction - Connecting Math Concepts, DISTAR Arithmetic I/II, Corrective Mathematics		
Becker & Gersten (1982)		insufficient match
Bereiter & Kurland (1981-1982)		pretest equivalence not established
Brent & DiObilda (1993)		no accounting for pretest scores
Mac Iver, Kemper, & Stringfield (2003)		insufficient data
Merrell (1996)		no adequate control group
Meyer (1984)		insufficient data
Vreeland, Vail, Bradley, Buetow, Cipriano, Green, Henshaw, & Huth (1994)		insufficient match
Wellington, J. (1994).		inadequate outcome measure
Wilson & Sindelar (1991)		duration < 12 weeks
Mastery Learning		
Burke (1980)		duration < 12 weeks
Cabezon (1984)		no accounting for pretest differences
Chan, Cole, & Cahill (1988)		duration < 12 weeks
Earnheart (1989)		duration < 12 weeks
Gallagher (1991)		no adequate control group
Long (1991)		> 1/2 std dev apart at pretest
CGI		

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Fennema, Carpenter, Franke, Levi, Jacobs, & Empson (1996)		no adequate control group
Villasenor & Kepner (1993)		insufficient match

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Consistency Management & Cooperative Discipline		
Freiberg, Stein, & Huang (1995)		subset of another study
Freiberg, Huzinec, & Borders (2006)		no adequate control group (artificial control group)
Project SEED		
Chadbourne & Webster (1996)		treatment confounded with other factors
Webster (1998).		not enough information, no pretest information
Webster & Chadbourn (1989).		treatment confounded with other factors
Webster & Chadbourn (1990).		treatment confounded with other factors
Webster & Chadbourn (1992).		treatment confounded with other factors
Cooperative Learning		
Al-Halal (2001)		duration < 12 weeks
Bosfield (2004)		insufficient data
De Russe (1999)		no adequate control group
Gabbert, Johnson, & Johnson (1986)		duration < 12 weeks
Gilbert-Macmillan (1983)		duration < 12 weeks
Goldberg (1989) (cooperative learning - TGT)		inadequate outcome measure
Hallmark (1994)		duration < 12 weeks
Johnson, Johnson, & Scott (1978)		duration < 12 weeks
Johnson (1985) (Groups of Four)		> 1/2 std dev apart at pretest
Lucker, Rosenfield, Sikes, & Aronson (1976). (Jigsaw)		duration < 12 weeks
Madden & Slavin (1983)		duration < 12 weeks
Martin (1986) (cooperative learning - TGT)		duration < 12 weeks
Morgan (1994)		duration < 12 weeks
Nattiv (1994) (cooperative learning)		duration < 12 weeks
Peterson, Janicki, & Swing (1981) (small-group instruction)		duration < 12 weeks
Swing and Peterson (1982) (small-group instruction)		duration < 12 weeks
Tieso (2005)		duration < 12 weeks
Zuber (1992)		duration < 12 weeks
Reciprocal Peer Tutoring		
Fantuzzo, Davis, & Ginsburg (1995)		duration < 12 weeks
Fantuzzo, King, & Heller, L.R. (1992)		inadequate control group
Fantuzzo, Polite, and Grayson (1990)		uneven attrition, duration < 12 weeks
Ginsburg-Block (1998)		duration < 12 weeks
Ginsburg-Block & Fantuzzo (1997)		duration < 12 weeks
Heller & Fantuzzo (1993)		test inherent to treatment
Pigott, Fantuzzo, & Clement (1986)		duration < 12 weeks
Curriculum-Based Measurement		

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Allinder & Oats (1997)		no control group , inadequate outcome measure
Fuchs, Fuchs, & Hamlett (1989)		pretest differences too large
Fuchs, Fuchs, Hamlett, Phillips, & Bentz (1994)		measure inherent to treatment
Fuchs, Fuchs, Hamlett, & Stecker (1991)		measure inherent to treatment
Clarke and Shinn (2004)		no adequate control group
Stecker and Fuchs (2000)		no adequate control group
Tsuei (2005)		inadequate comparison group
<u>Schema-Based Instruction</u>		
Fuchs, Fuchs, Finelli, Courey, & Hamlett (2004)		inadequate outcome measure
Fuchs, Fuchs, Finelli, Courey, Hamlett, Sones, and Hope (2006)		inadequate outcome measure
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, & Jancek (2003)		inadequate outcome measure
Fuchs, Fuchs, Prentice, Hamlett, Finelli, & Courey (2004)		inadequate outcome measure
Jitendra and Hoff (1996)		no control group
Jitendra, Griffin, McGoey, Bhat, & Riley (1998)		duration < 12 weeks
<u>Other</u>		
Ai (2002)		no adequate control group
Beirne-Smith (1991) (peer tutoring)		duration < 12 weeks
Burkhouse, Loftus, Sadowski, & Buzad (2003) (Thinking Math professional development)		no adequate control group
Burton (2005)		pretest equivalence not established
Campbell, Rowan, & Cheng (1995) (Project IMPACT)		inadequate outcome measure
Cardelle-Elawar (1990) (Metacognition)		duration < 12 weeks
Cardelle-Elawar (1994) (Metacognition)		inadequate outcome measure
Cobb, Wood, Yackel, Nicholls, Wheatley, Trigatti, & Perlwitz (1991) (Problem-Centered Instructional Approach)		pretest equivalence not established
Craig & Cairo (2005) (QUILT)		no adequate control group
ERIA (2003) (*Strength in Numbers)*		no adequate control group
Fischer (1990) (Part-Part Whole Curriculum)		duration < 12 weeks
Follmer (2001)		duration < 12 weeks
Fuchs, Fuchs, Hamlett, & Appleton (2002)		inadequate outcome measure
Fuchs, Fuchs, Karns, Hamlett, & Katzaroff (1999) (performance-assessment-driven instruction)		inadequate outcome measure
Fuchs, Fuchs, Karns, Hamlett, Katzaroff, & Dutka (1997) (task-focused goals treatment)		inadequate outcome measure
Fuchs, Fuchs, & Prentice (2004) (problem-solving treatment)		inadequate outcome measure
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter (2003) (self-regulated learning strategies)		inadequate outcome measure

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, and Jancek (2003) (explicitly teaching for transfer)		inadequate outcome measure
Fueyo and Bushell (1998) (number line procedures and peer tutoring)		duration < 12 weeks
Ginsburg-Block & Fantuzzo (1998) (NCTM standards-based intervention)		duration < 12 weeks
Griffin (2001) (Number Worlds)		inadequate outcome measure
Hickey, Moore, & Pellegrino (2001) (Jaspers)		inadequate outcome measures
Hiebert and Wearne (1993)		inadequate outcome measure
Hohn & Frey (2002) (SOLVED)		duration < 12 weeks
Hooper (1992)		duration < 12 weeks
Mason & Good (1993) (MMP, Two-Group and Whole-Class teaching)		measure inherent to treatment, no controlling for pretests
Miller (1991) (ability grouping)		pretest equivalence not established
Phares (1997) (ability grouping)		no adequate control group
Pratton & Hales (1986) (active participation)		duration < 12 weeks
Sharpley, Irvine, & Sharpley (1983) (cross-age tutoring)		duration < 12 weeks
Sloan (1993) (direct instruction)		pretest equivalence not established
Stallings (1985)		insufficient information
Stallings & Krasavage (1986) (Madeline Hunter Model)		pretest equivalence not established
Stallings, Robbins, Presbey, & Scott (1986) (Madeline Hunter Model)		insufficient information
White (1996) (TIPS Math)		duration < 12 weeks
Yager. Johnson, Johnson, & Snider (1986) (cooperative learning with group processing)		duration < 12 weeks

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.

Appendix 2
Table of Abbreviations

ANOVA- Analysis of Variance
ANCOVA- Analysis of Covariance
CAI- Computer Assisted Instruction
CAT- California Achievement Test
CCC- Computer Curriculum Corporation
CGI- Cognitively Guided Instruction
CMC- Connecting Math Concepts
CMCD- Consistency Management-Cooperative Discipline
CRCT- Criterion –Referenced Competency Test (Georgia)
CTBS- Comprehensive Test of Basic Skills
CWPT- Classwide Peer Tutoring
ERIC- Education Resources Information Center
HLM- Hierarchical Linear Modeling
ILS- Integrated learning system
ISTEP- Indiana Statewide Testing for Educational Progress
ITBS- Iowa Test of Basic Skills
MAT- Metropolitan Achievement Test
MCAS- Massachusetts Comprehensive Assessment System
MMP- Missouri Mathematics Project
NALT- Northwest Achievement Level Test
NAT- National Achievement Test
NCTM- National Council of Teachers of Mathematics
NRC- National Research Council
NSF- National Science Foundation
PALS- Peer Assisted Learning Strategies
SAT- Stanford Achievement Test
SES- Socio-economic status
SESAT-Stanford Early School Achievement Test
SRA- Science Research Associates
SRI- Scholastic Reading Inventory
STAD- Student Teams Achievement Division
STL- Student Team Learning
TAAS- Texas Assessment of Academic Skills
TAI- Team Assisted Individualization
TCAP- Tennessee Comprehensive Achievement Test
TERC- Technical Education Research Centers

The Best Evidence Encyclopedia is a free web site created by the Johns Hopkins University School of Education's Center for Data-Driven Reform in Education (CDDRE) under funding from the Institute of Education Sciences, U.S. Department of Education.